Variable outcome and methylation status according to CEBPA mutant type in double-mutated acute myeloid leukemia patients and the possible implications for treatment

EUROPEAN HEMATOLOGY ASSOCIATION
haematologica
Journal of the European Hematology Association

Variable outcome and methylation status according to CEBPA mutant type in double-mutated acute myeloid leukemia patients and the possible implications for treatment

by Dima El-Sharkawi, Duncan Sproul, Christopher G Allen, Andrew Feber, Melissa Wright, Robert K Hills, David C Linch, and Rosemary E Gale

Haematologica 2017 [Epub ahead of print]

Citation: El-Sharkawi D, Sproul D, Allen CG, Feber A, Wright M, Hills RK, Linch DC, and Gale RE. Variable outcome and methylation status according to CEBPA mutant type in double-mutated acute myeloid leukemia patients and the possible implications for treatment.
Haematologica. 2017; 102:xxx
doi:10.3324/haematol.2017.173096

Variable outcome and methylation status according to *CEBPA* mutant type in double-mutated acute myeloid leukemia patients and the possible implications for treatment

Dima El-Sharkawi,[1] Duncan Sproul,[2] Christopher G Allen,[1] Andrew Feber,[3] Melissa Wright,[4] Robert K Hills,[4] David C Linch[1] and Rosemary E Gale[1]

[1]Department of Haematology, UCL Cancer Institute, London, UK
[2]MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK
[3]Medical Genomics, UCL Cancer Institute, London, UK
[4]Centre for Trials Research, Cardiff University, Cardiff, UK

Running title: Variable outcome and methylation in *CEBPA*^{DM} AML

Corresponding author: Rosemary E Gale,
Department of Haematology, UCL Cancer Institute,
72 Huntley Street, London WC1E 6DD, United Kingdom
Tel: +44-20-7679-6232
Fax: +44-20-7679-6222
e-mail: rosemary.gale@ucl.ac.uk

Abstract word count: 249
Main text word count: 4000
Figures and Tables: 8
Supplemental files: 1

Trial registration: ISRCTN17833622 and ISRCTN17161961

Acknowledgments

The authors would like to thank Kerra Pearce at UCL Genomics for technical assistance.

Funding

This work was supported by Leukaemia and Lymphoma Research, now called Bloodwise, and the UK Medical Research Council, and was undertaken at UCL, which receives a proportion of funding from the Department of Health's NIHR Biomedical Research Centre's funding scheme.

ABSTRACT

Although *CEBPA* double-mutated (*CEBPA*^{DM}) acute myeloid leukemia is considered to be favorable-risk disease, relapse remains a major cause of treatment failure. Most *CEBPA*^{DM} patients have a classic biallelic mutant combination with an N-terminal mutation leading to production of p30 protein plus a C-terminal loss-of-function in-frame indel mutation (*CEBPA*$^{Classic\text{-}DM}$), but approximately one-third of cases have one or more non-classic mutations, with diverse combinations reported, and there is little information on the consequences of such mutants. We evaluated outcome in a cohort of 104 *CEBPA*^{DM} patients, 79 *CEBPA*$^{Classic\text{-}DM}$ and 25 with non-classic mutants, and found that the latter may have poorer survival (5-year overall survival 64% versus 46%, P=0.05), particularly post-relapse (41% versus 0%, P=0.02). However, for this analysis, all non-classic cases were grouped together, irrespective of mutant combination. As *CEBPA*^{DM} cases have been reported to be hypermethylated, we used methylation profiling to assess whether this could segregate the different mutants. We developed a *CEBPA*$^{Classic\text{-}DM}$ methylation signature from a preliminary cohort of 10 *CEBPA*^{DM} (including 8 *CEBPA*$^{Classic\text{-}DM}$) and 30 *CEBPA* wild-type (*CEBPA*^{WT}) samples, and independently validated the signature in 17 *CEBPA*$^{Classic\text{-}DM}$ cases. Assessment of the signature in 16 *CEBPA*^{DM} cases with different non-classic mutant combinations showed that only 31% had a methylation profile equivalent to *CEBPA*$^{Classic\text{-}DM}$ whereas for 69% the profile was either intermediate between *CEBPA*$^{Classic\text{-}DM}$ and *CEBPA*^{WT} or equivalent to *CEBPA*^{WT}. These results suggest that *CEBPA*^{DM} cases with non-classic mutants may be functionally different from those with *CEBPA*$^{Classic\text{-}DM}$ mutants and should not automatically be included in the same prognostic group.
(AML12 is registered under ISRCTN17833622 and AML15 under ISRCTN17161961).

Introduction

The *CEBPA* gene encodes CCAAT/enhancer binding protein-α (C/EBPα), a basic leucine zipper (bZIP) transcription factor that is essential for hematopoietic stem cell regulation and myeloid development.[1,2] The gene is mutated in approximately 8% of acute myeloid leukemia (AML) patients with intermediate-risk cytogenetics, and presence of biallelic double mutations (*CEBPA*^{DM}) in the absence of a *FLT3* internal tandem duplication (*FLT3*^{ITD}) is associated with a favorable prognosis.[3-9] In the current risk-adapted therapy strategies for AML,[10-12] these patients are classified as good-risk and therefore not usually considered for consolidation of first remission by allogeneic transplantation.[11,13]

Mutations occur throughout the single exon gene but predominate at the N and C termini.[14] N-terminal mutations are nearly always frameshift or nonsense mutations causing increased translation from an internal ATG start site and production of a truncated p30 protein that retains the same reading-frame as the full-length p42 protein but lacks the first transactivation domain (TAD1). The most common C-terminal mutations are in-frame indels in the bZIP DNA binding domain (DBD) or leucine zipper domain (LZD) that lead to loss of the ability to bind to DNA or dimerize, classified here as C-terminal loss-of-function (C-LOF). However, many other mutations have also been reported, including missense mutations in the DBD or LZD, missense mutations and in-frame indels in the mid-region, and frameshift or nonsense mutations in the mid-region or C-terminus. Some also present as homozygous alterations due to chromosome 19 uniparental disomy.[15,16] The most common combination of mutations is an N-terminal frameshift on one allele plus a C-terminal in-frame indel on the other allele that together are predicted to lead to complete loss of normal p42 C/EBPα activity,[14] hereafter called the classic *CEBPA*^{DM} combination (*CEBPA*$^{Classic\text{-}DM}$). This combination was identified in 204 of 305 *CEBPA*^{DM} cases (67%) with defined mutants reported in six studies containing ≥20 *CEBPA*^{DM} cases.[3,6,8,17-19] The remaining 101 cases had multiple different mutant combinations with diverse consequences; 54 (18% of total *CEBPA*^{DM}) would be predicted to produce just p30 due to the presence of a mid-region or C-terminal truncation, 12 (4%) would only produce a classic C-LOF protein, 19 (6%) p30 plus a C-terminal missense mutant, 5 (2%) just a C-missense mutant and 11 (4%) other mutant combinations.

Understanding the consequence of the different types of *CEBPA*^{DM} mutants is needed as this may impact on clinical outcome, but there is limited information available on the specific mutations. *CEBPA*^{DM} cases have gene expression profiles that are significantly different from single-mutated (*CEBPA*^{SM}) and wild-type (*CEBPA*^{WT}) cases.[3,7] However, recent data suggests that *CEBPA*^{DM} cases with non-classic combinations may not always cluster with other *CEBPA*^{DM} cases,[7,20] with three of seven such cases classified as negative for the *CEBPA*^{DM} expression profile in one report.[20] Furthermore, genotype stratification according to expression profiling may be confounded by *CEBPA*^{WT} cases with completely silenced *CEBPA* expression (*CEBPA*^{SIL}) due to methylation of the *CEBPA* promoter as these cases cluster together with *CEBPA*^{DM} cases.[21] *CEBPA*^{DM} cases also form distinct epigenetic clusters with a markedly hypermethylated profile,[22,23] but although *CEBPA*^{SIL} cases also have a hypermethylated profile, this segregates from the *CEBPA*^{DM} cluster and, of note, they appear to be associated with a biologically distinct subtype of AML with a poor prognosis.[21,22,24] Potential methylation differences according to the underlying combination of *CEBPA* mutants, however, have not been reported.

In order to investigate potential differences between *CEBPA*^{DM} mutant combinations, we evaluated clinical outcome in 104 *CEBPA*^{DM} cases, 79 *CEBPA*$^{Classic\text{-}DM}$ and 25 with non-classic mutants, and observed that the non-classic cases may have a lower overall survival (OS). However, the number of cases was relatively small and all cases with a non-classic mutant were included in this group, irrespective of the mutant type. We therefore investigated whether methylation profiling of samples from double-mutated patients could assist in segregating the different mutant combinations.

Methods

Patient cohorts

The patients investigated were younger adults entered into the UK MRC AML10, AML12 and AML15 trials. Informed patient consent was obtained in accordance with the Declaration of Helsinki, and ethical approval for tissue use from the Wales Research Ethics Committee 3. Clinical outcome was evaluated in 104 *CEBPA*^{DM} patients, all <60 years of age, and methylation profiling was performed on 135 patients, 132 (98%) of them <60 years of age (Supplementary Figure S1). Selected tests were performed on a further 82 samples with specific cytogenetic and molecular abnormalities.

Therapy, clinical end points and statistical methods

Details of clinical protocols, end points and statistical methods are defined in the Supplementary Material.

Methylation arrays and data processing

DNA was bisulfite-converted using the EZ DNA Methylation-Gold Kit (Zymo Research, California, USA) and random samples checked by methylation-specific PCR to ensure efficient conversion (see Supplementary Material). Methylation profiling was performed using the Illumina Infinium HumanMethylation27 (n=40, cohort 1) and HumanMethylation450 (n=95, cohort 2) BeadChip arrays (Illumina Inc, California, USA). Details of data processing are given in the Supplemental Material. Derived beta values were expressed as the percentage methylation at a given CpG probe. Selected CpG sites were further analyzed using pyrosequencing assays (see Supplementary Material).

Unsupervised analysis of patient methylation profiles

In a given sample, probes were defined as methylated if the beta value was >0.3, unmethylated if ≤0.3.[25] Samples were clustered based on their beta values at probes displaying significant variation (variable probes) as previously defined.[25,26] They were defined as variable if methylated in ≥1 sample(s) plus unmethylated in ≥1 sample(s). CpG islands (CGIs) were located as previously defined;[27] their methylation levels were derived by calculating the mean beta value of probes at these locations. Hierarchical clustering was performed using the Euclidian distance of beta values and Ward algorithm in R.

Derivation and analysis of *CEBPA* signature

A methylation signature of *CEBPA* genotype was derived from variably methylated CpGs in cohort 1. Signature CpGs were selected as the top 25 ranked probes based on the mean rank of *P* values of Wilcoxon tests and the absolute median difference in beta value between $CEBPA^{\text{Classic-DM}}$ and $CEBPA^{\text{WT}}$ samples. Probes were included in the analysis if >90% of samples and ≥2 $CEBPA^{\text{Classic-DM}}$ and ≥2 $CEBPA^{\text{WT}}$ samples had observable data. Methylation signatures of $CEBPA^{\text{Classic-DM}}$ and $CEBPA^{\text{WT}}$ samples were then defined as the median beta values observed at these 25 probes. Samples were

scored relative to these signatures by calculating the Euclidian distance between the signatures and their profiles at signature probes.

Mutant *CEBPA* level and confirmation of biallelic status

Mutant *CEBPA* level was quantified as previously described[6] or approximated from the sequence chromatogram (average height of ≥5 peaks). Monoallelic/biallelic status was investigated by sequencing clones derived from full-length *CEBPA* amplicons as previously described.[6]

RESULTS

Clinical outcome in *CEBPA*^{DM} cases according to mutant type

Of the 104 *CEBPA*^{DM} cases evaluated, 79 had classic and 25 had non-classic mutants (Table 1, Supplementary Table S1). The latter included 4 patients predicted to produce p30 plus a C-terminal missense mutant, 8 with combinations predicted to produce just p30, and 13 with different combinations predicted to produce just a C-LOF protein. There was no difference in the baseline characteristics between the classic and non-classic cases, including white cell count, sex, WHO performance status and AML type (de novo or secondary), although the non-classic cases were older (median 35 versus 47 years, P=0.001) (Supplementary Table S2). All classic cases had intermediate-risk cytogenetics compared to 89% of the non-classic cases (P=0.05). There was no significant difference in the incidence of *FLT3*^{ITD}, *NPM1*, *IDH1*, *WT1*, *TET2* and *GATA2* mutations between the groups, although the non-classic cases had more *IDH2* mutations (1% versus 20%, P=0.003; all *IDH2*^{R140Q}) and *DNMT3A* mutations (3% versus 16%, P=0.03). Median follow-up was 9.5 years (range, 0.1-22 years). Neither the proportion of transplanted patients, the type of transplant nor the stage of transplantation differed between the groups (Supplementary Table S2).

CEBPA$^{Classic\text{-}DM}$ cases had a slightly higher but statistically non-significantly different complete remission (CR) rate to non-classic cases (95% versus 88%, P=0.2) (Table 2). In univariate analysis, there was no significant difference in relapse-free survival or relapse rate (RR). However, there was a trend towards a lower OS in the non-classic cases (64% versus 46% at 5 years, P=0.05) (Figure 1A), which was largely due to the worse outcome post-relapse in the non-classic cases. The proportion of relapsing

patients who achieved a second remission did not differ between the groups (61% versus 55%), but 5-year survival post-relapse was 41% versus 0% (P=0.02) (Figure 1B), and 59% versus 0% from second remission (P=0.004) (Figure 1C). The survival differences were not, however, statistically different in multivariate analysis (Table 2), but this is not surprising given the small group sizes. Although the significantly higher proportion of *DNMT3A*-mutated cases in the non-classic group could have contributed to their worse outcome, as both these mutations and $FLT3^{ITD}$ adversely impact on the favorable outcome associated with $CEBPA^{DM}$ AML,[6,28] the trend towards a lower OS in the non-classic cases was still present when patients with these mutations were excluded (70% versus 53%; Hazard ratio 1.95, 95% confidence intervals 0.95-4.01; P=0.08) (Figure 1D).

Development of a $CEBPA^{Classic\text{-}DM}$ methylation signature

The clinical evaluation grouped all non-classic cases together in order to obtain an adequate number of patients for analysis, but this cohort therefore included patients with many mutant combinations predicted to have differing functional consequences. In order to explore potential methods for discriminating between these combinations, we investigated the impact of mutant type on methylation profiles. A preliminary cohort of samples from 40 normal karyotype (NK) $FLT3^{WT}NPM1^{WT}$ patients were investigated using the Illumina Infinium HumanMethylation27 array; 10 were $CEBPA^{DM}$ and 30 were $CEBPA^{WT}$ (Methylation cohort 1, Table 1), approximating the relative proportions of such cases in NK $FLT3^{WT}NPM1^{WT}$ patients in our earlier study.[6] The array data was validated by pyrosequencing assays at four differentially methylated CpG sites (Supplementary Figure S2). Most CpG sites analyzed showed little variation in methylation levels across the whole cohort, but unsupervised cluster analysis according to levels of the most variably methylated probes revealed two main clusters. All 10 $CEBPA^{DM}$ samples, including two non-classic cases, fell in the cluster of 16 samples with significantly higher levels of mean CGI methylation (Figure 2A), and the mean level of CGI methylation was significantly different between $CEBPA^{Classic\text{-}DM}$ and $CEBPA^{WT}$ samples (Figure 2B).

A supervised approach was then used to create $CEBPA^{Classic\text{-}DM}$ and $CEBPA^{WT}$ methylation signatures based on the 25 most differentially methylated sites between the $CEBPA^{Classic\text{-}}$

^{DM} and *CEBPA*^{WT} samples (Figure 2C, Supplementary Table S3). Two distance scores were calculated for each sample based on the Euclidian distance between their methylation levels at these signature probes and the median profile of the *CEBPA*$^{Classic\text{-}DM}$ and *CEBPA*^{WT} samples. This confirmed that, when assessed according to their distance scores, the *CEBPA*$^{Classic\text{-}DM}$ samples formed a distinct group that clearly separated from the *CEBPA*^{WT} samples (Figure 3A).

The *CEBPA*$^{Classic\text{-}DM}$ methylation signature was validated with samples from a further 17 *CEBPA*$^{Classic\text{-}DM}$ and 26 *CEBPA*^{WT} cases (Methylation cohort 2, Table 1) using the HumanMethylation450 array. Sixteen of the 17 *CEBPA*$^{Classic\text{-}DM}$ cases (94%) fell in the same cluster in unsupervised analysis, with a relatively more hypermethylated profile, and all *CEBPA*^{WT} cases fell in the hypomethylated cluster (Supplementary Figure S3). Using supervised analysis according to the derived *CEBPA*$^{Classic\text{-}DM}$ and *CEBPA*^{WT} signatures, the same 16 *CEBPA*^{DM} cases had a methylation profile that was closest to the *CEBPA*$^{Classic\text{-}DM}$ signature with a wide difference between the two signatures (Figures 3B, 4), indicating that their profile was equivalent to the *CEBPA*$^{Classic\text{-}DM}$ cases in cohort 1. Further analysis of the one *CEBPA*$^{Classic\text{-}DM}$ case that fell closer to the *CEBPA*^{WT} group indicated that the mutations were biallelic but only approximately half of the cells in the sample carried mutations, mean mutant level 28% for the pair, which was the lowest mean level of all 25 *CEBPA*$^{Classic\text{-}DM}$ cases (median 44%; range, 28%-50%). The methylation profile of this case could therefore have been affected by the presence of a significant proportion of non-leukemic cells and it was excluded from further analyses. Using the distance scores (mean±2SD) for *CEBPA*$^{Classic\text{-}DM}$ cases in cohort 1, tests showed that these scores classified *CEBPA*$^{Classic\text{-}DM}$ and *CEBPA*^{WT} genotypes in the second cohort with 95% accuracy, 88% sensitivity and 100% specificity. The distance scores of the 24 *CEBPA*$^{Classic\text{-}DM}$ cases from the combined cohorts were then used to define a *CEBPA*$^{Classic\text{-}DM}$ quadrant that segregated all *CEBPA*$^{Classic\text{-}DM}$ cases from *CEBPA*^{WT} cases (Figure 3C).

Investigation of the *CEBPA*$^{Classic\text{-}DM}$ methylation signature in other good-risk patients

In order to examine whether the *CEBPA*$^{Classic\text{-}DM}$ methylation signature could simply reflect 'good-risk' disease or be due to a lack of C/EBPα expression, methylation levels at three differentially methylated CpG sites from the signature were quantified by

pyrosequencing using samples from 21 patients with inv(16) and 19 with t(8;21), both associated with downregulated C/EBPα expression,[29-31] and 42 with *NPM1*^{MUT}*FLT3*^{WT}. *KHNYN* and *VAMP5* were more hypermethylated in the *CEBPA*$^{Classic\text{-}DM}$ signature, *LY9* more hypomethylated. Each subgroup differed significantly from the 24 *CEBPA*$^{Classic\text{-}DM}$ cases at two of the three sites (Supplementary Figure S4). A composite methylation score was calculated for each sample; it was statistically significantly different from the *CEBPA*$^{Classic\text{-}DM}$ cases for all three subgroups (Figure 5), indicating that the methylation profile was a distinct feature of the mutant proteins that lead to total loss of normal C/EBPα function rather than absence of C/EBPα per se.

Investigation of non-classic *CEBPA*^{DM} and *CEBPA*^{SM} mutants

Having established that *CEBPA*$^{Classic\text{-}DM}$ cases have a methylation profile that is distinct from *CEBPA*^{WT} cases, profiles of 14 *CEBPA*^{DM} cases with a variety of different non-classic combinations were investigated using the HumanMethylation450 array (Methylation cohort 2, Table 1). On unsupervised analysis, nine (64%) were hypermethylated and five (36%) hypomethylated (Supplementary Figure S3). To assess their impact on the *CEBPA* methylation signatures, these 14 cases and the two non-classic cases from the initial cohort were considered according to the predicted functional consequence of the combination.

Six cases were predicted to produce just p30 protein, with a classic N plus null mutant (mid-region or C-terminal frameshift/nonsense) combination. Only one fulfilled the *CEBPA*$^{Classic\text{-}DM}$ criteria; four fell outside this quadrant but were still distinct from *CEBPA*^{WT} cases, and one grouped with the *CEBPA*^{WT} cases (Figure 6A). Three cases were predicted to just give rise to classic C-LOF proteins without p30 (two homozygous classic C, one compound heterozygous with a classic C plus C-frameshift combination). Two fulfilled the *CEBPA*$^{Classic\text{-}DM}$ criteria and one was intermediate between *CEBPA*$^{Classic\text{-}DM}$ and *CEBPA*^{WT}. The seven cases with missense mutations were also highly variable. Three had a classic N plus C-missense combination; none fulfilled the *CEBPA*$^{Classic\text{-}DM}$ criteria, two were intermediate between the signatures and one grouped with *CEBPA*^{WT} cases. Mutant levels were indicative of 80% or more mutated cells in all three cases; two were biallelic by cloning but no full-length amplicons could be obtained in one case. Similarly, only one of the three homozygous C-missense cases fell in the *CEBPA*$^{Classic\text{-}DM}$

quadrant, the other two were equivalent to *CEBPA*^{WT} cases. The remaining case was compound heterozygous with a classic C and C-missense combination, and did fulfil the *CEBPA*$^{Classic\text{-}DM}$ criteria.

These results suggest that the functional consequence of double-mutated cases producing at least one non-classic mutant protein can be highly variable and difficult to predict. Of note, when outcome was assessed in the non-classic cases according to their methylation profile, there was a suggestion that those that fulfilled the *CEBPA*$^{Classic\text{-}DM}$ methylation criteria were less likely to relapse, with only 1 of 4 cases (25%) relapsing compared to 7 of 10 cases (70%) that fell outside the *CEBPA*$^{Classic\text{-}DM}$ quadrant (5-year RR 25% versus 54% respectively). These numbers were too small for meaningful statistical analysis, and the results do not necessarily indicate a causal link between the methylation pattern and outcome. They do suggest, however, that the methylation pattern could be a useful biomarker, and thereby act as a surrogate for response and selection of therapy.

Methylation cohort 2 also included 38 *CEBPA*^{SM} cases with a wide range of classic and non-classic mutants (Table 1). On unsupervised analysis, 31 (82%) were in the hypomethylated cluster, with no apparent segregation between the *CEBPA*^{SM} and *CEBPA*^{WT} cases (Supplementary Figure S3). On supervised analysis, all except two cases were equivalent to *CEBPA*^{WT} cases, with no obvious grouping according to the type of mutant (Figures 4, 6B). The remaining two cases fell in or close to the *CEBPA*$^{Classic\text{-}DM}$ quadrant; both had a classic C mutation. It is possible that the *CEBPA*^{WT} allele had been silenced in these cases, but RNA samples were not available to check this.

Discussion

Molecular genotyping is increasingly used to risk stratify patients with AML, but clinical application of this information needs to be accurate and robust for optimal patient therapy. This is particularly important for good-risk patients such as those with biallelic *CEBPA* mutations, where the current recommendation is not to proceed to transplantation in first remission, as for some patients this could lead to under-treatment. Identifying those who are at greater risk of relapse and poorer survival may therefore guide patient management. Very limited information is available on the impact

of the different *CEBPA* mutations, with all *CEBPA*^{DM} cases currently being classified as good-risk, irrespective of the underlying mutant. However, mutations identified in approximately one-third of *CEBPA*^{DM} patients do not conform to the classic combination of N-terminal frameshift or nonsense mutation plus C-terminal in-frame indel. As many of the other mutations are unique and of unknown functional consequence, determining their significance is challenging, particularly in view of the multiple functions attributed to C/EBPα.[1,2]

Although we had access to a database of 2162 patients with known *CEBPA* genotype and available clinical data from three consecutive UK MRC trials of younger adult patients with AML, 67% with intermediate-risk cytogenetics, we were still able to evaluate long-term outcome in a cohort of only 79 *CEBPA*$^{Classic\text{-}DM}$ cases and 25 *CEBPA*^{DM} cases with at least one non-classic mutant. The results suggested that the non-classic cases may have a poorer outcome; in particular, none of these cases survived after relapse. These results may be influenced by differences in the coincident mutations in the two groups. The majority of recurrent mutations with known prognostic significance in AML are uncommon in *CEBPA*^{DM} cases,[32] and for many of them their impact in this subgroup is therefore not well defined. Both *FLT3*^{ITD} and *DNMT3A* mutations adversely impact on the favorable outcome of *CEBPA*^{DM} AML.[6,28] but even when patients with these mutations were excluded from the analysis, OS for the non-classic cases was still worse. The incidence of *GATA2* mutations was non-significantly lower in the non-classic cases, but although one study reported that they are associated with better OS,[33] most studies, including our own, observed no difference.[34,35] There was a non-significantly higher incidence of *TET2* mutations in the non-classic cases, but their impact is unclear, with only one study reported specifically for *CEBPA*^{DM} AML that showed a worse OS but not event-free survival in *TET2*^{MUT} cases.[33] Clearly, many more cases would need to be analyzed in order to take into account coincident mutations other than *FLT3*^{ITD} and *DNMT3A* mutations.

A wide range of non-classic mutations was observed and, for the outcome analysis, sufficient patient numbers could only be obtained by grouping all the non-classic patients together, which precluded evaluation of specific mutant combinations. We therefore sought alternative methods of assessment and, as *CEBPA*^{DM} patients have a

distinct hypermethylated profile,[22,23] investigated whether genome-wide methylation profiling could provide information on the more broad-spectrum functional consequence of different mutants. We confirmed the relatively hypermethylated profile associated with a *CEBPA*^{DM} genotype in a preliminary cohort and then derived methylation signatures for *CEBPA*$^{Classic\text{-}DM}$ and *CEBPA*^{WT} cases using the 25 most differentially methylated CpG sites. These were validated in an independent cohort of samples, with 16 of the 17 *CEBPA*$^{Classic\text{-}DM}$ cases studied clearly separating from the *CEBPA*^{WT} cases. The remaining case contained a significant proportion of non-mutated cells, which provided indirect evidence that the signatures reflected *CEBPA* genotype. The signatures were a distinct feature of mutant C/EBPα and could not simply be attributed to a lack of C/EBPα, as they were not replicated in samples from patients with core-binding factor leukemias that are associated with down-regulation of *CEBPA* expression. Presence of at least a minimal level of C/EBPα activity is thought to be necessary for the development of leukemia as *Cebpa*$^{-/-}$ mice with totally absent C/EBPα accumulate immature myeloid progenitors but do not develop AML,[36,37] and AML patients have not been reported with mutations leading to complete absence of C/EBPα. These results therefore suggest that it is the functionally aberrant C/EBPα protein that underlies the hypermethylated profile detected in the *CEBPA*^{DM} cases.

The *CEBPA*$^{Classic\text{-}DM}$ cases provided a framework for assessing the methylation profiles of mutant combinations with at least one non-classic mutant. Only 31% had a methylation profile equivalent to *CEBPA*$^{Classic\text{-}DM}$, 25% were equivalent to *CEBPA*^{WT}, and 44% were intermediate between the two. Similar heterogeneity has been reported for gene expression profiles of non-classic cases, with three of seven such cases segregating from cases with classical mutants.[20] This variability is not surprising considering the diversity of the mutant combinations. Although the p30 isoform is thought to play an important role in allowing commitment of the leukemic stem cell to the myeloid lineage,[38] the mechanism by which it promotes AML is not clearly defined. It has a lower affinity for some C/EBP sites than p42 and induces multiple genes that are not affected by p42,[39-41] and this may have influenced the methylation profile. Knock/in mice expressing just N-terminal mutant developed leukemia but more slowly than the N+C combination.[38,42] This presumably reflects the additional influence of an aberrant C-terminally mutated protein that might not bind to DNA but can still bind to other C/EBP interacting proteins

such as PU.1 and the SWI/SNF complex. Classic C-terminal mutants are associated with hyperproliferation due to loss of cell-cycle regulation and a block in myeloid differentiation.[43] Although knock/in mice with classic C mutant alone do develop leukemia, it is with slower latency than the N+C and mutant N mice.[42] Since these mutants may still bind and potentially sequester other interacting factors, it has been suggested that this could limit the ability of other C/EBPs to rescue the effect of the aberrant C/EBPα,[2] as shown for C/EBPβ in the C/EBPα-deficient situation.[44,45] This more global cellular impact of the C-terminal mutants may have a greater consequence for signaling events downstream of C/EBPα, which may therefore be reflected in the methylation profile.

The methylation profiles did not group according to the predicted functional consequence of the mutant, whether N- or C-terminally mutated. For example, considerable variability was observed for the cases with a C-missense mutation. Of the three classic N plus C-missense combinations assayed, one case had a methylation profile equivalent to *CEBPA*^{WT} and two cases had intermediate profiles. Two homozygous C-missense cases grouped with *CEBPA*^{WT} cases, whereas another homozygous C-missense case grouped with the *CEBPA*$^{Classic\text{-}DM}$ cases. In the 51 cases documented with C-missense mutations in the COSMIC database (http://cancer.sanger.ac.uk/cosmic), most mutations are unique, reported in one (n=27, 53%) or two (n=9, 18%) patients, and only two residues (R297 and R300) are recorded as being variably mutated in five patients. Critical amino acids at the bZIP/DNA interface have been identified from the C/EBPα crystal structure, but many additional hydrogen bonds and van der Waals contacts are implicated in the stabilization of these interactions,[46] and predicting the functional consequence of these mutants is therefore difficult.

From a clinical perspective, risk management requires evaluation of all the available information, and the data presented here suggests that *CEBPA*^{DM} patients with a non-classic mutation should not automatically be included in the same favorable-risk prognostic group as *CEBPA*$^{Classic\text{-}DM}$ cases, and that it might be appropriate to consider them for allogeneic transplantation in first remission. Ultimately, this can only be proven by analysis of clinical trial outcomes and, with the increasing availability of large

data sets using targeted next-generation sequencing panels, such analysis may be feasible in the future. This will also facilitate a better understanding of the mutational background of classic and non-classic *CEBPA*^{DM} cases and whether there are differences that impact on their prognosis. Our studies also raise the possibility that methylation profiling may identify those non-classic cases that behave in a similar manner to classic mutants, although we cannot directly attribute a causal link between the methylation pattern and chemosensitivity, and further studies are required before this is introduced into clinical practice.

Conflict of interest

The authors have no conflicts of interest to declare.

References

1. Friedman AD. C/EBPalpha in normal and malignant myelopoiesis. Int J Hematol. 2015;101(4):330-341.
2. Ohlsson E, Schuster MB, Hasemann M, Porse BT. The multifaceted functions of C/EBPalpha in normal and malignant haematopoiesis. Leukemia. 2016;30(4):767-775.
3. Wouters BJ, Lowenberg B, Erpelinck-Verschueren CA, van Putten WL, Valk PJ, Delwel R. Double *CEBPA* mutations, but not single *CEBPA* mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. Blood. 2009;113(13):3088-3091.
4. Dufour A, Schneider F, Metzeler KH, et al. Acute myeloid leukemia with biallelic *CEBPA* gene mutations and normal karyotype represents a distinct genetic entity associated with a favorable clinical outcome. J Clin Oncol. 2010;28(4):570-577.
5. Pabst T, Eyholzer M, Fos J, Mueller BU. Heterogeneity within AML with CEBPA mutations; only CEBPA double mutations, but not single CEBPA mutations are associated with favourable prognosis. Br J Cancer. 2009;100(8):1343-1346.
6. Green CL, Koo KK, Hills RK, Burnett AK, Linch DC, Gale RE. Prognostic significance of *CEBPA* mutations in a large cohort of younger adult patients with acute myeloid leukemia: impact of double *CEBPA* mutations and the interaction with FLT3 and NPM1 mutations. J Clin Oncol. 2010;28(16):2739-2747.
7. Taskesen E, Bullinger L, Corbacioglu A, et al. Prognostic impact, concurrent genetic mutations, and gene expression features of AML with *CEBPA* mutations in a cohort of 1182 cytogenetically normal AML patients: further evidence for *CEBPA* double mutant AML as a distinctive disease entity. Blood. 2011;117(8):2469-2475.
8. Fasan A, Haferlach C, Alpermann T, et al. The role of different genetic subtypes of *CEBPA* mutated AML. Leukemia. 2014;28(4):794-803.
9. Li HY, Deng DH, Huang Y, et al. Favorable prognosis of biallelic *CEBPA* gene mutations in acute myeloid leukemia patients: a meta-analysis. Eur J Haematol. 2015;94(5):439-448.
10. Dohner H, Estey EH, Amadori S, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood. 2010;115(3):453-474.

11. O'Donnell MR, Abboud CN, Altman J, et al. Acute myeloid leukemia. J Natl Comp Cancer Network. 2012;10(8):984-1021.
12. Ofran Y, Rowe JM. Genetic profiling in acute myeloid leukaemia--where are we and what is its role in patient management. Br J Haematol. 2013;160(3):303-320.
13. Cornelissen JJ, Gratwohl A, Schlenk RF, et al. The European LeukemiaNet AML Working Party consensus statement on allogeneic HSCT for patients with AML in remission: an integrated-risk adapted approach. Nat Rev Clin Oncol. 2012;9(10):579-590.
14. Nerlov C. C/EBPalpha mutations in acute myeloid leukaemias. Nat Rev Cancer. 2004;4(5):394-400.
15. Fitzgibbon J, Smith LL, Raghavan M, et al. Association between acquired uniparental disomy and homozygous gene mutation in acute myeloid leukemias. Cancer Res. 2005;65(20):9152-9154.
16. Wouters BJ, Sanders MA, Lugthart S, et al. Segmental uniparental disomy as a recurrent mechanism for homozygous *CEBPA* mutations in acute myeloid leukemia. Leukemia. 2007;21(11):2382-2384.
17. Shih LY, Liang DC, Huang CF, et al. AML patients with CEBPalpha mutations mostly retain identical mutant patterns but frequently change in allelic distribution at relapse: a comparative analysis on paired diagnosis and relapse samples. Leukemia. 2006;20(4):604-609.
18. Benthaus T, Schneider F, Mellert G, et al. Rapid and sensitive screening for *CEBPA* mutations in acute myeloid leukaemia. Br J Haematol. 2008;143(2):230-239.
19. Behdad A, Weigelin HC, Elenitoba-Johnson KS, Betz BL. A clinical grade sequencing-based assay for *CEBPA* mutation testing: report of a large series of myeloid neoplasms. J Mol Diag. 2015;17(1):76-84.
20. Lavallee VP, Krosl J, Lemieux S, et al. Chemo-genomic interrogation of *CEBPA* mutated AML reveals recurrent *CSF3R* mutations and subgroup sensitivity to JAK inhibitors. Blood. 2016;127(24):3054-3061.
21. Wouters BJ, Jorda MA, Keeshan K, et al. Distinct gene expression profiles of acute myeloid/T-lymphoid leukemia with silenced *CEBPA* and mutations in *NOTCH1*. Blood. 2007;110(10):3706-3714.
22. Figueroa ME, Lugthart S, Li Y, et al. DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia. Cancer Cell. 2010;17(1):13-27.

23. Cancer Genome Atlas Research Network, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. New Engl J Med. 2013;368(22):2059-2074.
24. Figueroa ME, Wouters BJ, Skrabanek L, et al. Genome-wide epigenetic analysis delineates a biologically distinct immature acute leukemia with myeloid/T-lymphoid features. Blood. 2009;113(12):2795-2804.
25. Sproul D, Kitchen RR, Nestor CE, et al. Tissue of origin determines cancer-associated CpG island promoter hypermethylation patterns. Genome Biol. 2012;13(10):R84.
26. Sproul D, Nestor C, Culley J, et al. Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. Proc Natl Acad Sci USA. 2011;108(11):4364-4369.
27. Illingworth RS, Gruenewald-Schneider U, Webb S, et al. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Gen. 2010;6(9):e1001134.
28. Gale RE, Lamb K, Allen C, El-Sharkawi D, Stowe C, Jenkinson S, et al. Simpson's Paradox and the Impact of Different *DNMT3A* Mutations on Outcome in Younger Adults With Acute Myeloid Leukemia. J Clin Oncol. 2015;33(18):2072-2083.
29. Pabst T, Mueller BU, Harakawa N, et al. AML1-ETO downregulates the granulocytic differentiation factor C/EBPalpha in t(8;21) myeloid leukemia. Nat Med. 2001;7(4):444-451.
30. Ptasinska A, Assi SA, Martinez-Soria N, et al. Identification of a dynamic core transcriptional network in t(8;21) AML that regulates differentiation block and self-renewal. Cell Rep. 2014;8(6):1974-1988.
31. Helbling D, Mueller BU, Timchenko NA, et al. CBFB-SMMHC is correlated with increased calreticulin expression and suppresses the granulocytic differentiation factor CEBPA in AML with inv(16). Blood. 2005;106(4):1369-1375.
32. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. New Engl J Med. 2016;374(23):2209-2221.
33. Grossmann V, Haferlach C, Nadarajah N, et al. *CEBPA* double-mutated acute myeloid leukaemia harbours concomitant molecular mutations in 76.8% of cases with *TET2* and *GATA2* alterations impacting prognosis. Br J Haematol. 2013;161(5):649-658.

34. Green CL, Tawana K, Hills RK, et al. *GATA2* mutations in sporadic and familial acute myeloid leukaemia patients with *CEBPA* mutations. Br J Haematol. 2013;161(5):701-705.
35. Theis F, Corbacioglu A, Gaidzik VI, et al. Clinical impact of *GATA2* mutations in acute myeloid leukemia patients harboring *CEBPA* mutations: a study of the AML study group. Leukemia. 2016;30(11):2248-2250.
31. Heath V, Suh HC, Holman M, et al. C/EBPalpha deficiency results in hyperproliferation of hematopoietic progenitor cells and disrupts macrophage development in vitro and in vivo. Blood. 2004;104(6):1639-1647.
32. Zhang P, Iwasaki-Arai J, Iwasaki H, et al. Enhancement of hematopoietic stem cell repopulating capacity and self-renewal in the absence of the transcription factor C/EBP alpha. Immunity. 2004;21(6):853-863.
33. Kirstetter P, Schuster MB, Bereshchenko O, et al. Modeling of C/EBPalpha mutant acute myeloid leukemia reveals a common expression signature of committed myeloid leukemia-initiating cells. Cancer Cell. 2008;13(4):299-310.
34. Cleaves R, Wang QF, Friedman AD. C/EBPalphap30, a myeloid leukemia oncoprotein, limits G-CSF receptor expression but not terminal granulopoiesis via site-selective inhibition of C/EBP DNA binding. Oncogene. 2004;23(3):716-725.
35. Geletu M, Balkhi MY, Peer Zada AA, et al. Target proteins of C/EBPalphap30 in AML: C/EBPalphap30 enhances sumoylation of C/EBPalphap42 via up-regulation of Ubc9. Blood. 2007;110(9):3301-3309.
36. Wang C, Chen X, Wang Y, Gong J, Hu G. C/EBPalphap30 plays transcriptional regulatory roles distinct from C/EBPalphap42. Cell Res. 2007;17(4):374-383.
37. Bereshchenko O, Mancini E, Moore S, et al. Hematopoietic stem cell expansion precedes the generation of committed myeloid leukemia-initiating cells in C/EBPalpha mutant AML. Cancer Cell. 2009;16(5):390-400.
38. Kato N, Kitaura J, Doki N, et al. Two types of C/EBPalpha mutations play distinct but collaborative roles in leukemogenesis: lessons from clinical data and BMT models. Blood. 2011;117(1):221-233.
39. Jones LC, Lin ML, Chen SS, et al. Expression of C/EBPbeta from the C/ebpalpha gene locus is sufficient for normal hematopoiesis in vivo. Blood. 2002;99(6):2032-2036.
40. Hirai H, Zhang P, Dayaram T, et al. C/EBPbeta is required for 'emergency' granulopoiesis. Nat Immunol. 2006;7(7):732-739.

41. Miller M, Shuman JD, Sebastian T, Dauter Z, Johnson PF. Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. J Biol Chem. 2003;278(17):15178-15184.

Table 1. ***CEBPA*** **genotype of investigated cohorts**

	CEBPA genotype*	n	Predicted functional consequence
Clinical cohort (n=104)	*CEBPA* $^{\text{Classic-DM}}$	79	p30 + C-LOF
	Non-classic *CEBPA* $^{\text{DM}}$		
	Classic N + C-missense	4	p30 + C-LOF
	Classic N + mid-frameshift	5	p30
	Classic N + C-frameshift	2	p30
	Homozygous classic N	1	p30
	Homozygous classic C	4	C-LOF
	Homozygous C-missense	3	C-LOF
	Classic C + C-frameshift	4	C-LOF
	Classic C + C-missense	1	C-LOF
	Mid-frameshift + C-missense	1	C-LOF
Methylation Cohort 1 (n=40)	*CEBPA* $^{\text{Classic-DM}}$	8	p30 + C-LOF
	Non-classic *CEBPA* $^{\text{DM}}$		
	Classic C + C-frameshift	1	C-LOF
	Homozygous C-missense	1	C-LOF
	CEBPA $^{\text{WT}}$	30	WT
Methylation Cohort 2 (n=95)	*CEBPA* $^{\text{Classic-DM}}$	17	p30 + C-LOF
	Non-classic *CEBPA* $^{\text{DM}}$		
	Classic N + C-missense	3	p30 + C-LOF
	Classic N + mid-frameshift	5	p30
	Classic N + C-frameshift	1	p30
	Homozygous classic C	2	C-LOF
	Homozygous C-missense	2	C-LOF
	Classic C + C-missense	1	C-LOF
	CEBPA $^{\text{SM}}$		
	Classic N	9	p30 + WT
	Classic C	5	C-LOF + WT
	Mid-indel	3	UNK + WT
	Mid-frameshift	9	Null**+ WT
	Mid-missense	5	UNK + WT
	C-frameshift	2	Null** + WT
	C-missense	5	C-LOF + WT
	CEBPA $^{\text{WT}}$	26	WT

*Details of the specific mutations are given in Supplementary Table S1.
**Mid-region or C-terminal mutants with a truncating frameshift or nonsense mutation
Abbreviations: C, C-terminal mutation; C-LOF, C-terminal loss-of-function; DM, double mutant; indel, in-frame insertion and/or deletion; N, N-terminal mutation; SM, single mutant; UNK, unknown; WT, wild-type

Table 2. Outcome according to *CEBPA*^{DM} mutant combination

Outcome	Classic *CEBPA*^{DM} (n=79)	Non-classic *CEBPA*^{DM} (n=25)	*CEBPA*$^{Classic\text{-}DM}$ vs non-classic *CEBPA*^{DM} OR or HR (95% CI), *P*-value	
			Univariate	Multivariate*
CR	95%	88%**	3.20 (0.53-19.47); P=0.2	Not evaluable***
5 yr OS	64%	46%	1.84 (1.01-3.37), *P*=0.05	1.44 (0.74-2.79), *P* =0.3
5 yr RFS	49%	45%	1.15 (0.61-2.16), *P*=0.7	1.03 (0.52-2.06), *P* =0.9
5 yr RR	41%	77%	1.32 (0.66-2.66), *P*=0.4	1.09 (0.47-2.55), *P* =0.8

*Adjusted for age, white blood cell count, WHO performance status, type of leukemia, sex, *FLT3* and *DNMT3A* genotype.

**Remission status was missing for one patient

***Insufficient events for analysis

Abbreviations: CI, confidence intervals; CR, complete remission; DM, double mutant; HR, hazard ratio; OR, odds ratio; OS, overall survival; RFS, relapse-free survival; RR, relapse rate

Figure Legends

Figure 1. Clinical outcome in *CEBPA*^{DM} patients according to mutant combination. (A) Overall survival in the total cohort of 104 *CEBPA*^{DM} patients. (B) Survival post-relapse in the 39 patients that relapsed. (C) Survival from second remission in the 23 relapsed patients who achieved a second remission. (D) Overall survival in the total cohort excluding patients with *FLT3*^{ITD} and *DNMT3A* mutations.

Figure 2. Methylation profile of the preliminary cohort of 40 samples. (A) Unsupervised cluster analysis of the cohort. Each column represents a patient. Genotype is given in the upper panel. Samples were clustered based on their methylation levels at 7,679 variable probes, and the heatmap in the middle panel shows the variable CpG probes located within CpG islands (CGIs). The latter were used to calculate the mean % CpG methylation shown in the lower panel; red and blue bars indicate a predominantly hyper- or hypo-methylated profile respectively. (B) The mean level CGI methylation for *CEBPA*^{WT} and *CEBPA*$^{Classic\text{-}DM}$ samples. CGI methylation levels were calculated from all autosomal CGI probes. ***P<0.001 (Wilcoxon test). (C) Supervised cluster analysis showing the derived *CEBPA*$^{Classic\text{-}DM}$ signature on the left and the *CEBPA*^{WT} signature on the right. Samples are ordered according to their level of similarity to the *CEBPA*$^{Classic\text{-}DM}$ signature using Euclidian distance. The lower panel shows the distance scores indicating the distance from the *CEBPA*$^{Classic\text{-}DM}$ (green circles) and *CEBPA*^{WT} (white circles) signatures; the lower the y axis value, the more closely the sample matches that particular signature.

Figure 3. Analysis of the methylation profiles of the samples according to their distances from the derived *CEBPA*$^{Classic\text{-}DM}$ and *CEBPA*^{WT} methylation signatures. Difference between the distance scores (*CEBPA*$^{Classic\text{-}DM}$ - *CEBPA*^{WT}) compared with the distance from the *CEBPA*$^{Classic\text{-}DM}$ (MUT) signature of (A) the preliminary cohort and (B) the *CEBPA*$^{Classic\text{-}DM}$ and *CEBPA*^{WT} cases in the follow-up cohort. (C) Criteria derived for *CEBPA*$^{Classic\text{-}DM}$ using mean±2SD of the distance scores of all 24 *CEBPA*$^{Classic\text{-}DM}$ cases.

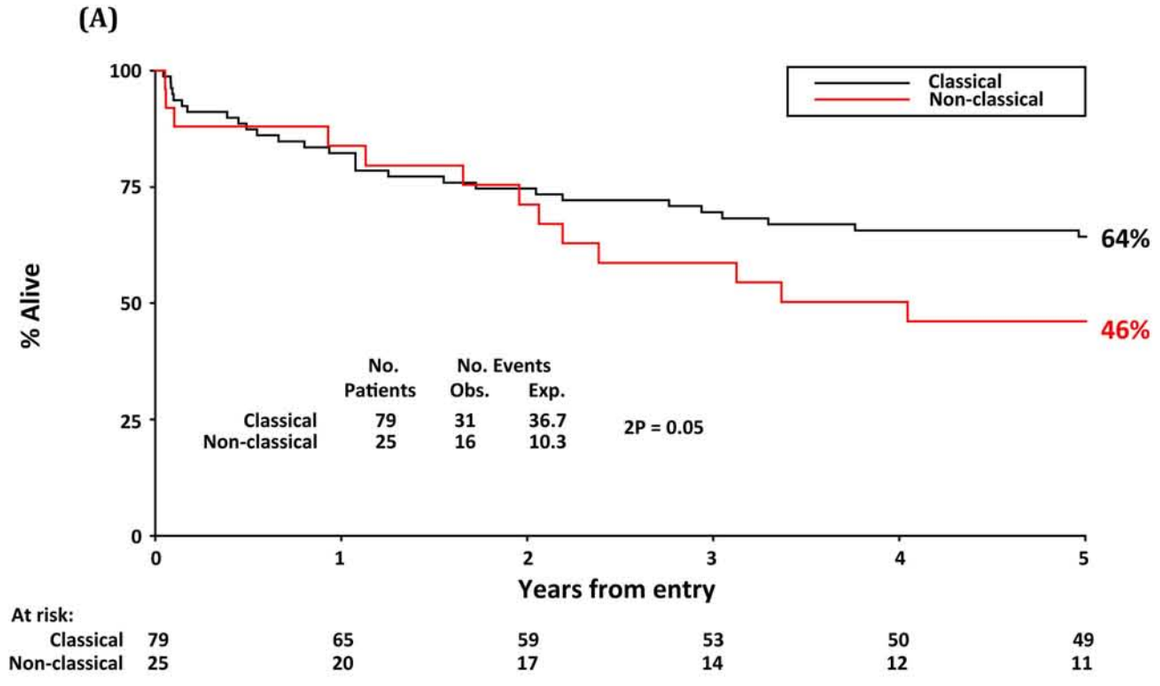
Figure 4. Analysis of the follow-up cohort of 95 samples. Supervised analysis according to the *CEBPA*$^{Classic\text{-}DM}$ and *CEBPA*^{WT} methylation signatures. Patient *CEBPA*

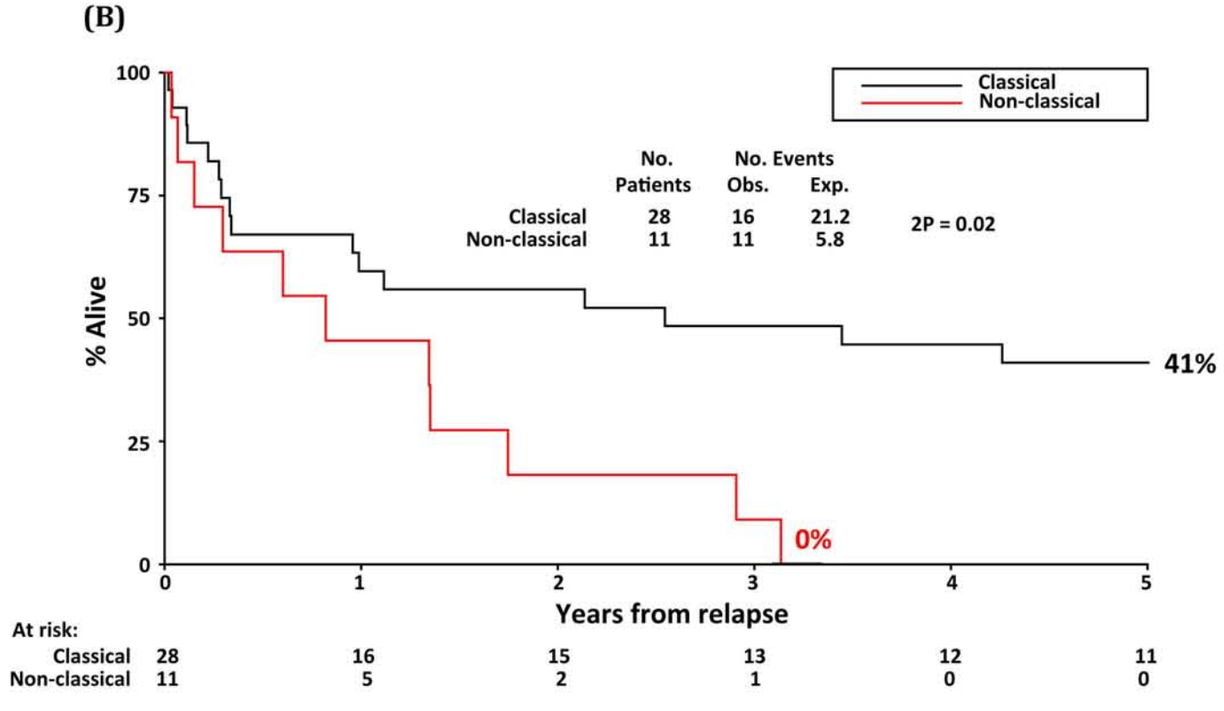
genotype is given above the heatmap. The lower panel shows the distance scores indicating the distance from the *CEBPA*$^{\text{Classic-DM}}$ (green circles) and *CEBPA*$^{\text{WT}}$ (white circles) signatures; the lower the y axis value, the more closely the sample matches that particular signature.

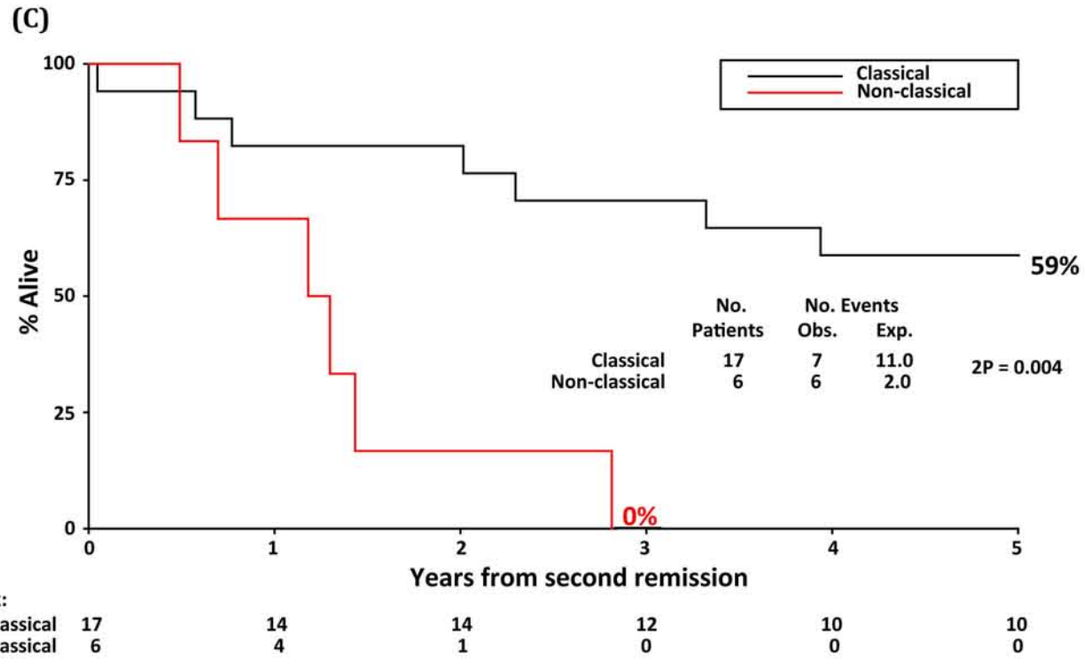
Figure 5. Methylation levels in other good-risk groups compared to the *CEBPA*$^{\text{Classic-DM}}$ cases. A composite methylation score was calculated by summing the difference between the % methylation for samples and the median for the 24 *CEBPA*$^{\text{Classic-DM}}$ cases (excluding the outlier) for three differentially methylated CpG sites, *KHNYN*, *VAMP5* and *LY9*. Mean values±95% confidence intervals are shown. The *CEBPA*$^{\text{Classic-DM}}$ and *CEBPA*$^{\text{WT}}$ results were beta values from the arrays; results for the three comparative groups were obtained by pyrosequencing. Significance refers to difference from the *CEBPA*$^{\text{Classic-DM}}$ group (*$P \leq 0.05$, **$P \leq 0.01$, ***$P \leq 0.001$).

Figure 6. Supervised analysis of *CEBPA*$^{\text{DM}}$ cases with non-classic mutations and *CEBPA*$^{\text{SM}}$ cases. Difference between the distance scores (*CEBPA*$^{\text{Classic-DM}}$ - *CEBPA*$^{\text{WT}}$) compared with the distance from the *CEBPA*$^{\text{Classic-DM}}$ (MUT) signature for (A) *CEBPA*$^{\text{DM}}$ cases with a non-classic mutant. (B) *CEBPA*$^{\text{SM}}$ cases.

Figure 1.


(A)
Classical
Non-classical
% Alive
100
75
50
25
0
64%
46%
No. Patients
No. Events Obs.
Exp.
Classical 79 31 36.7
Non-classical 25 16 10.3
2P = 0.05
0 1 2 3 4 5
Years from entry
At risk:
Classical 79 65 59 53 50 49
Non-classical 25 20 17 14 12 11


(B)
Classical
Non-classical
% Alive
100
75
50
25
0
41%
0%
No. Patients
No. Events Obs.
Exp.
Classical 28 16 21.2
Non-classical 11 11 5.8
2P = 0.02
0 1 2 3 4 5
Years from relapse
At risk:
Classical 28 16 15 13 12 11
Non-classical 11 5 2 1 0 0


(C)
Classical
Non-classical
% Alive
100
75
50
25
0
59%
0%
No. Patients
No. Events Obs.
Exp.
Classical 17 7 11.0
Non-classical 6 6 2.0
2P = 0.004
0 1 2 3 4 5
Years from second remission
At risk:
Classical 17 14 14 12 10 10
Non-classical 6 4 1 0 0 0

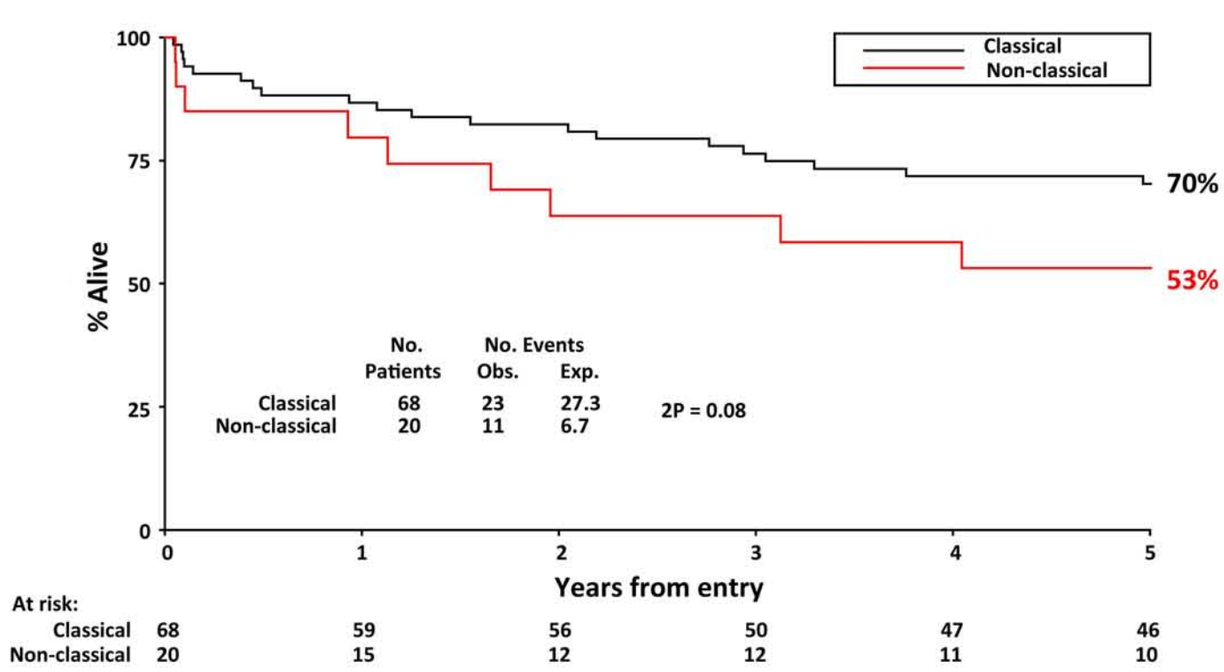

(D)
Classical
Non-classical
% Alive
100
75
50
25
0
70%
53%
No. Patients
No. Events Obs.
Exp.
Classical 68 23 27.3
Non-classical 20 11 6.7
2P = 0.08
0 1 2 3 4 5
Years from entry
At risk:
Classical 68 59 56 50 47 46
Non-classical 20 15 12 12 11 10

Figure 2.

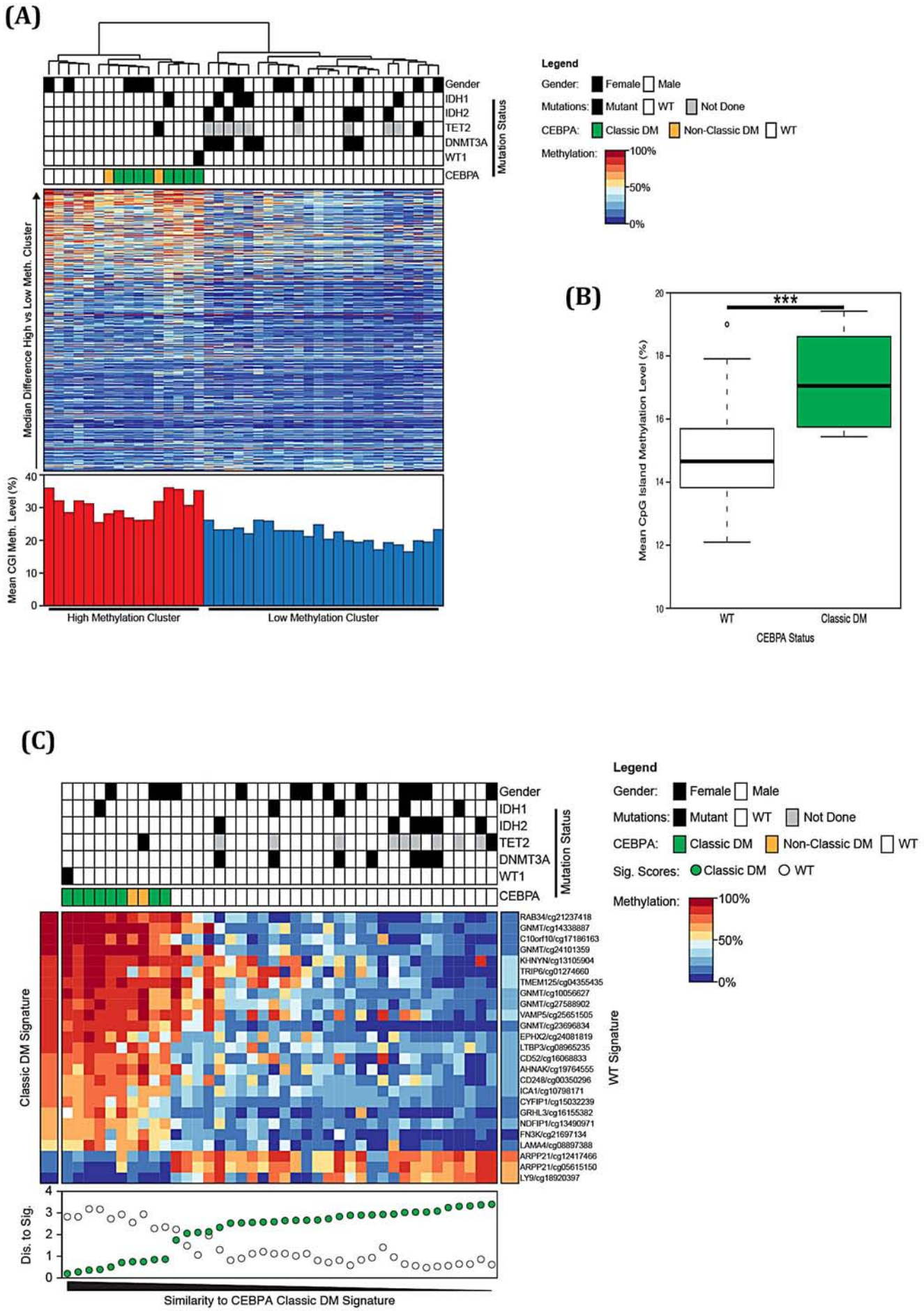

(A)
Gender
IDH1
IDH2
TET2
DNMT3A
WT1
CEBPA
Mutation Status
Legend
Gender: Female Male
Mutations: Mutant WT Not Done
CEBPA: Classic DM Non-Classic DM WT
Methylation: 100% 50% 0%
Median Difference High vs Low Meth. Cluster
Mean CGI Meth. Level (%)
High Methylation Cluster
Low Methylation Cluster
(B)
***
Mean CpG Island Methylation Level (%)
WT
Classic DM
CEBPA Status
(C)
Gender
IDH1
IDH2
TET2
DNMT3A
WT1
CEBPA
Mutation Status
Legend
Gender: Female Male
Mutations: Mutant WT Not Done
CEBPA: Classic DM Non-Classic DM WT
Sig. Scores: Classic DM WT
Methylation: 100% 50% 0%
RAB34/cg21237418
GNMT/cg14338887
C10orf10/cg17186163
GNMT/cg24101359
KIHNYN/cg13105904
TRIP6/cg01274660
TMEM125/cg04355435
GNMT/cg10056627
GNMT/cg27588902
VAMP5/cg25651505
GNMT/cg23696834
EPHX2/cg24081819
LTBP3/cg08965235
CD52/cg16068833
AHNAK/cg19764555
CD248/cg00350296
ICA1/cg10798171
CYFIP1/cg15032239
GRHL3/cg16155382
NDFIP1/cg13490971
FN3K/cg21697134
LAMA4/cg08897388
ARPP21/cg12417466
ARPP21/cg05615150
LY9/cg18920397
Classic DM Signature
WT Signature
Dis. to Sig.
Similarity to CEBPA Classic DM Signature

Figure 3.

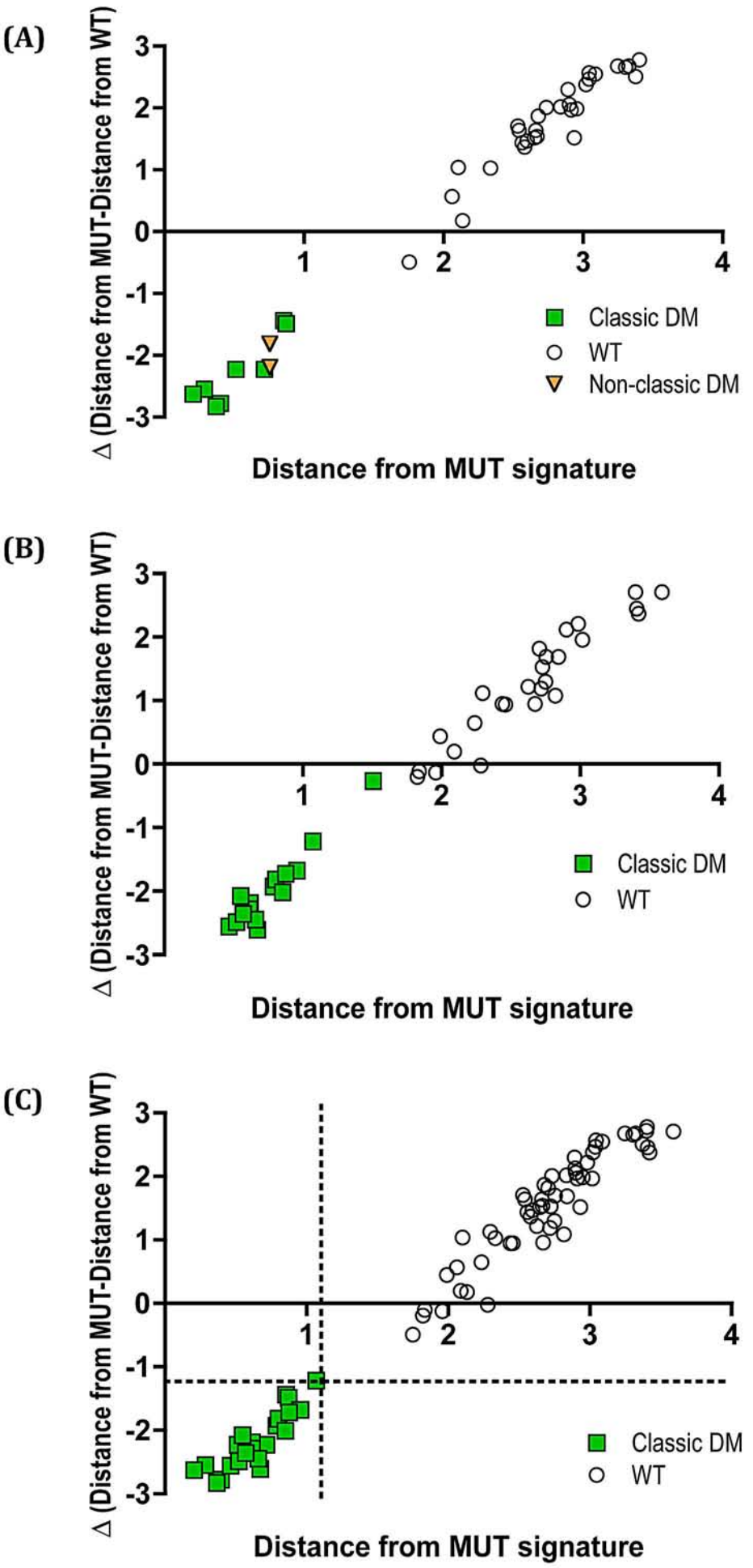
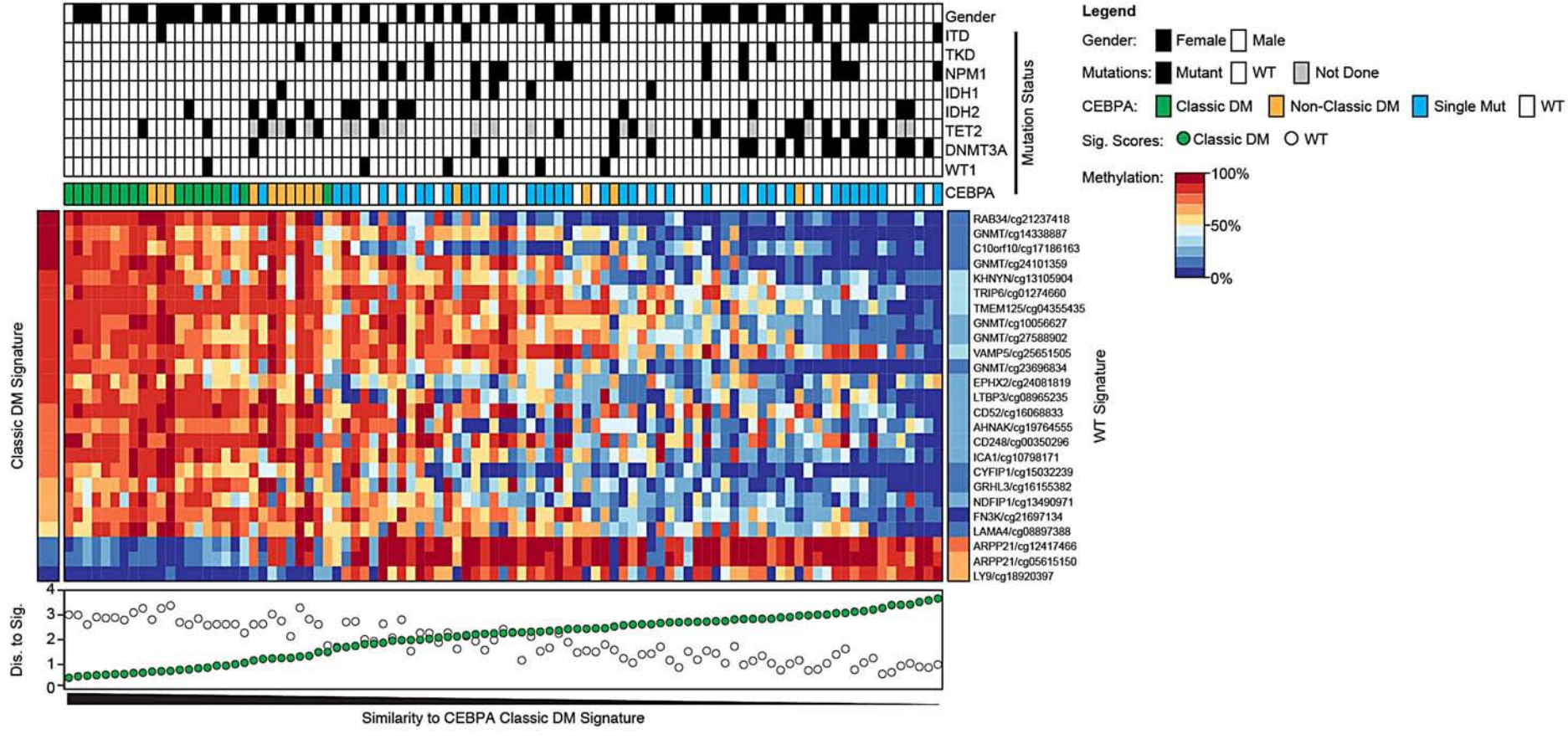

(A)
Δ (Distance from MUT-Distance from WT)
Distance from MUT signature
Classic DM
WT
Non-classic DM
(B)
Δ (Distance from MUT-Distance from WT)
Distance from MUT signature
Classic DM
WT
(C)
Δ (Distance from MUT-Distance from WT)
Distance from MUT signature
Classic DM
WT

Figure 4.


Gender
ITD
TKD
NPM1
IDH1
IDH2
TET2
DNMT3A
WT1
Mutation Status
CEBPA
Legend
Gender: Female Male
Mutations: Mutant WT Not Done
CEBPA: Classic DM Non-Classic DM Single Mut WT
Sig. Scores: Classic DM WT
Methylation: 100% 50% 0%
Classic DM Signature
RAB34/cg21237418
GNMT/cg14338887
C10orf10/cg17186163
GNMT/cg24101359
KHNYN/cg13105904
TRIP6/cg01274660
TMEM125/cg04355435
GNMT/cg10056627
GNMT/cg27588902
VAMP5/cg25651505
GNMT/cg23696834
EPHX2/cg24081819
LTBP3/cg08965235
CD52/cg16068833
AHNAK/cg19764555
CD248/cg00350296
ICA1/cg10798171
CYFIP1/cg15032239
GRHL3/cg16155382
NDFIP1/cg13490971
FN3K/cg21697134
LAMA4/cg08897388
ARPP21/cg12417466
ARPP21/cg05615150
LY9/cg18920397
WT Signature
Dis. to Sig.
4
3
2
1
0
Similarity to CEBPA Classic DM Signature

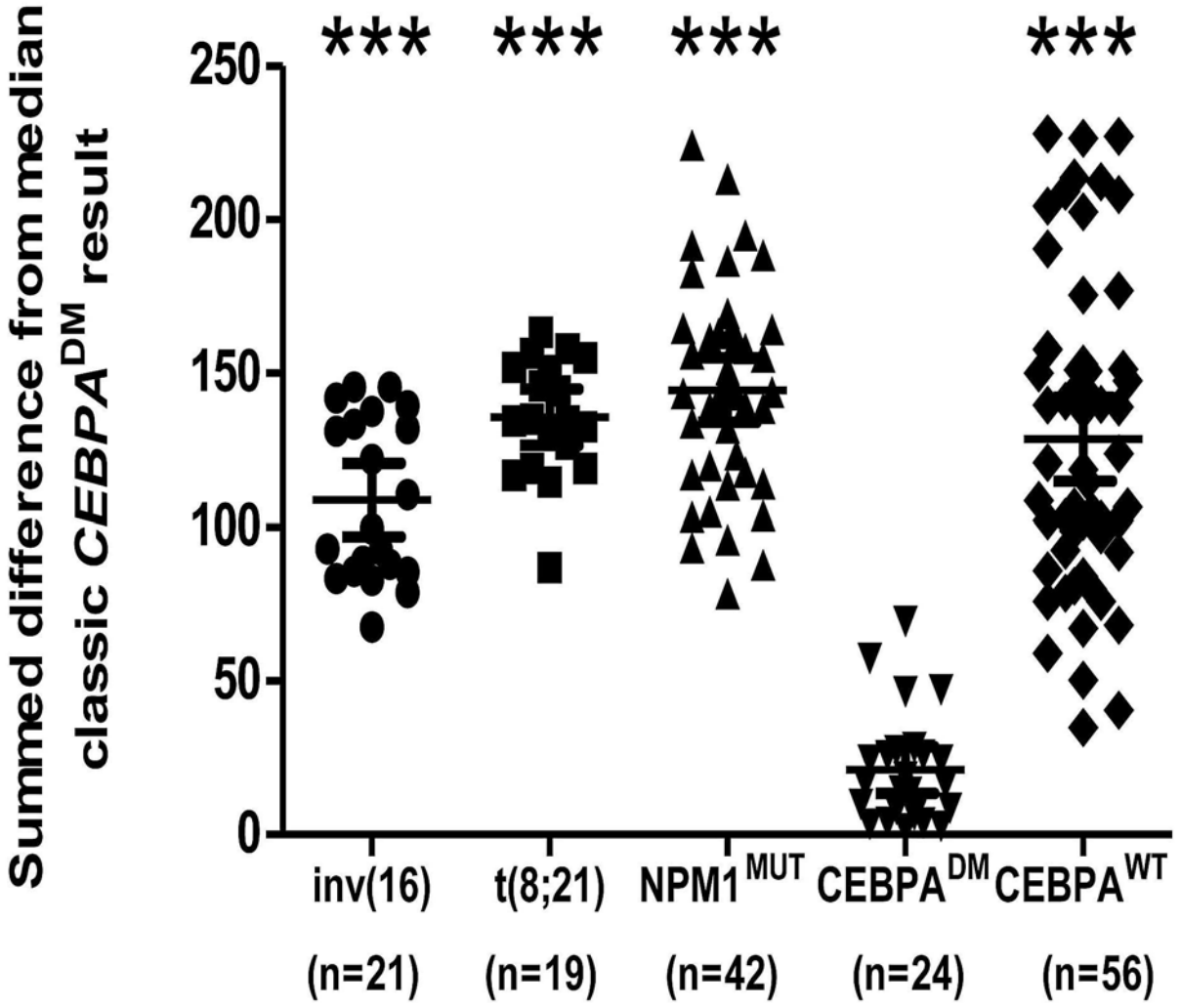
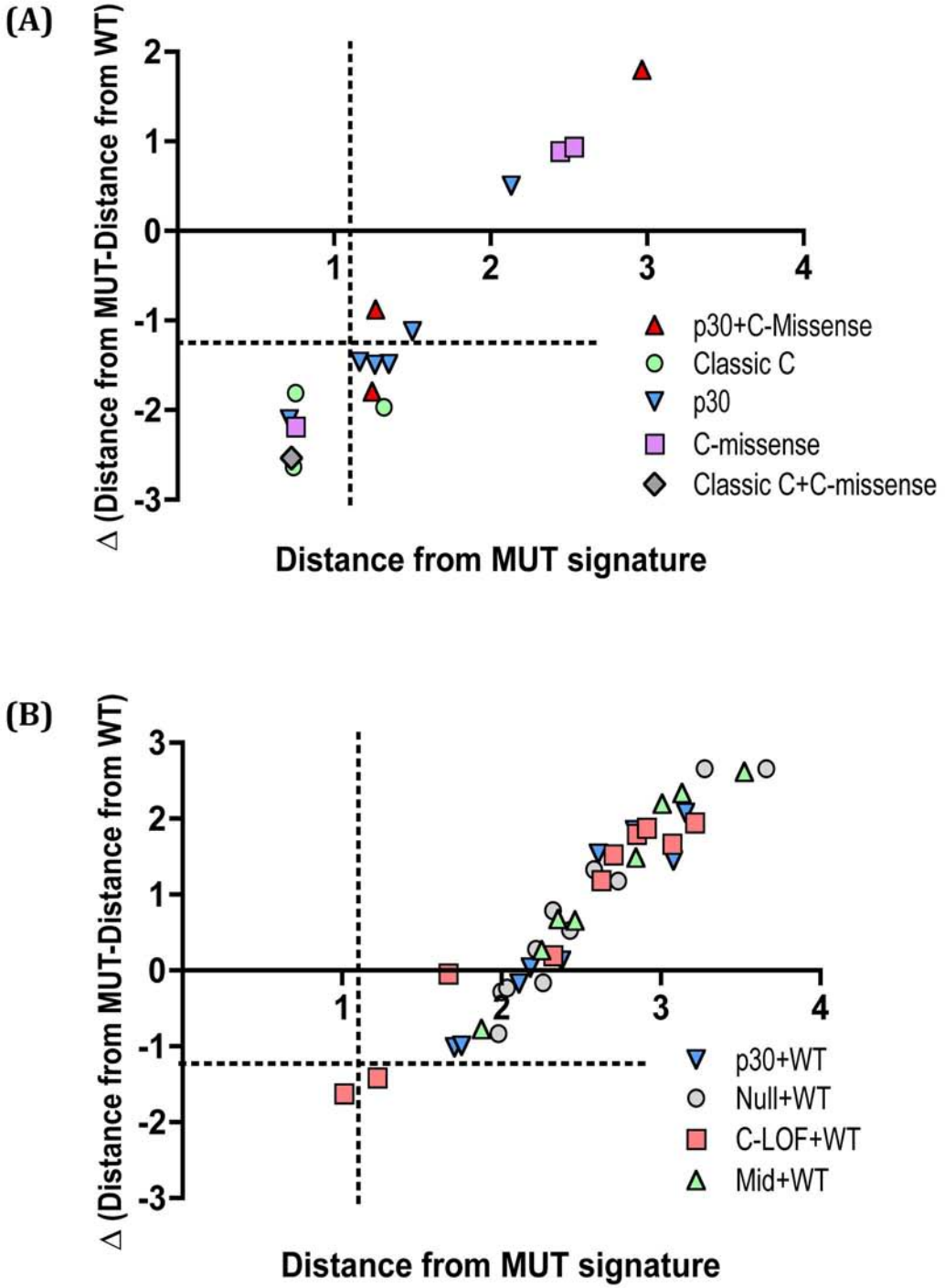

Summed difference from median classic CEBPA^DM result
250
200
150
100
50
0
***
***
***
***
inv(16)
t(8;21)
NPM1^MUT
CEBPA^DM
CEBPA^WT
(n=21)
(n=19)
(n=42)
(n=24)
(n=56)

Figure 5.

Figure 6.


(A)
Δ (Distance from MUT-Distance from WT)
Distance from MUT signature
p30+C-Missense
Classic C
p30
C-missense
Classic C+C-missense
(B)
Δ (Distance from MUT-Distance from WT)
Distance from MUT signature
p30+WT
Null+WT
C-LOF+WT
Mid+WT

SUPPLEMENTARY MATERIAL

Therapy, clinical end points and statistical methods

Details of the clinical protocols have been published elsewhere.[1-3] CR was defined as a normocellular bone marrow (BM) containing <5% blasts and showing evidence of normal maturation of other marrow elements. Persistence of myelodysplastic features did not preclude the diagnosis of CR. OS was the time from trial entry to death. For patients achieving CR, relapse-free survival (RFS) was the time from the date of first CR to an event (death in first CR or relapse) and relapse rate (RR) was the cumulative probability of relapse, censoring at death in CR.

Mantel-Haenszel and chi-squared tests were used to test for differences in demographic and clinical data by genotype. Kaplan-Meier curves were constructed for survival data and compared by means of the log-rank test. Surviving patients were censored on August 2010 for AML10 and AML12, March 2015 for AML15. Median follow-up for survival was 9.5 years (range, 0.1-22 years). Multivariable Cox models were used to analyze OS, RFS and RR, adjusting for age, white cell count, WHO performance status, secondary leukemia and trial, with additional variables of interest (*CEBPA* mutant type, *FLT3*^{ITD} and *DNMT3A* genotype). Models were fitted using forward selection, with variables added to the model if they had a *P* value, derived using the deviance statistic, of <0.05. Odds ratios (OR) or hazard ratios (HR) and 95% confidence intervals (CI) are quoted for endpoints. In all cases a ratio of <1 indicates benefit. All *P* values are two-tailed.

Bisulfite conversion and methylation-specific PCR to assess conversion efficiency

An aliquot of 350-500ng DNA was bisulfite-converted using the EZ DNA Methylation-Gold Kit (Zymo Research, California, USA) according to the manufacturer's instructions except that the incubation at 50^{o}C for 16 hours included a denaturation step at 95^{o}C for 30 seconds at the beginning of each hour. This periodic cycling has been reported to improve conversion efficiency.[4]

Several samples were randomly selected from each converted batch and subjected to two methylation-specific PCRs of the *HLA-B* gene, one using primers that would only

amplify bisulfite-converted DNA and the other using primers that would only amplify unconverted DNA (see Table S5). For each PCR, 1µl of converted DNA was added to 1x manufacturer's buffer, 3mM $MgCl_2$, 200µM dNTPs, 25pmols each primer and 1U GoTaq polymerase (Promega) in a total volume of 25µl. An initial denaturation step at 95°C for 6 minutes was followed by 36 cycles of PCR, each of 94°C for 30 seconds, annealing temperature for 30 seconds and 72°C for 30 seconds, with 2 cycles at annealing temperatures of 60°C, 59°C and 58°C, then 30 cycles at 57°C, followed by a final extension step of 72°C for 15 minutes. PCR products were run on a 2% agarose gel. Samples were considered to be successfully bisulfite-converted if they had a PCR product with the primers for the bisulfite-converted DNA but no product with the primers for the unconverted DNA.

Data processing for the methylation arrays

Data from the Illumina Infinium HumanMethylation27 and HumanMethylation450 BeadChip arrays (Illumina Inc, California, USA) were exported from Illumina's genome studio and subsequent analysis was performed using R. Beta values were used as a measure of the methylation level at a given CpG probe as derived from the intensity of the methylated (I_{meth}), and unmethylated (I_{unmeth}) allele probes (I_{meth} /(I_{meth} + I_{unmeth}). The data from 450K arrays was normalized to correct for systematic biases between the two types of probes present on these arrays as described.[5] Beta values were then filtered to remove unreliable data points based on the detection *P* value from the Infinium arrays (threshold 0.01) by setting their values to NA. Finally, probes displaying gender-specific biases were filtered out, defining them as those with Wilcoxon tests *P* values <0.05 between genders. This resulted in the exclusion of 1,362 probes from the first cohort (Infinium 27K array) and 56,356 probes from the follow-up cohort (Infinium 450K array). Beta values were converted to percentage methylation levels by multiplying by 100.

Pyrosequencing assays

Pyrosequencing assays to interrogate specific CpG sites were designed using the PyroMark Assay Design Software 2.0 (Qiagen, Crawley, UK). For each PCR, 25ng DNA was added to 1x manufacturer's buffer, 3.5mM $MgCl_2$, 200µM dNTPs, 0.2µM primers, one biotin-labeled and one unlabeled, and 1.25U GoTaq polymerase (Promega) in a total

volume of 20µl. The mixes were denatured at 95°C for 5 minutes and then subjected to 50 cycles of amplification using primers and annealing temperatures as specified in Table S5. Products were sequenced on a PyroMark MD system, (Qiagen) using PyroMark Q96 reagents and protocols and the appropriate primer as specified in Table S5. All samples were assayed in duplicate and the mean methylation level expressed as a percentage of the total alleles for a particular site. For each assay, titration curves were created using standards containing 0%, 10%, 25%, 50%, 75%, 90% and 100% methylated DNA, prepared from bisulfite-converted fully methylated and unmethylated DNA (Epitect Control DNA set, Qiagen), to check for accuracy and sensitivity.

REFERENCES

1. Hann IM, Stevens RF, Goldstone AH, et al. Randomized comparison of DAT versus ADE as induction chemotherapy in children and younger adults with acute myeloid leukemia. Results of the Medical Research Council's 10th AML trial (MRC AML10). Adult and Childhood Leukaemia Working Parties of the Medical Research Council. Blood. 1997;89(7):2311-2318.
2. Burnett AK, Hills RK, Milligan DW, et al. Attempts to optimize induction and consolidation treatment in acute myeloid leukemia: results of the MRC AML12 trial. J Clin Oncol. 2010;28(4):586-595.
3. Burnett AK, Russell NH, Hills RK, et al. Optimization of chemotherapy for younger patients with acute myeloid leukemia: results of the medical research council AML15 trial. J Clin Oncol. 2013;31(27):3360-3368.
4. Thirlwell C, Eymard M, Feber A, et al. Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. Methods. 2010;52(3):248-254.
5. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011;3(6):771-784.

SUPPLEMENTARY TABLES

Table S1. Full details of *CEBPA* mutations in the patients studied

Coh	Genotype	Mutant 1	Mutant 2	Type of mutation	Predicted proteins
Clin	Classic DM	p.H18fs	p.E309del	Classic N + C	p30 + C-LOF
		p.H18fs	p.Q312del		
		p.S21fs	p.N307_Q312del		
		p.S21fs	T310dup		
		p.P23fs	K304_Q305insL		
		p.P23fs	p.N307_E309delinsK		
		p.P23fs	p.L315_E316insT		
		p.H24fs	p.K304_Q305insL		
		p.H24fs	p.K304_R306dup		
		p.H24fs	p.V308dup		
		p.H24fs	p.Q311_Q312insL		
		p.H24fs	p.Q311_Q312insL		
		p.H24fs	p.K313Nins14		
		p.H24fs	p.K313_V314insQK		
		p.A29fs	p.L315dup		
		p.A30fs	p.Q312Hins15		
		p.F31fs	p.K313dup		
		p.F33fs	p.K326delins23		
		p.R35fs	p.E309dup		
		p.G36fs	p.R297_Q311dup		
		p.G36fs	p.D301_L315dup		
		p.G36fs	p.K304_Q312dup		
		p.G36fs	p.R306_N307insKQR		
		p.G36fs	p.E309delinsKTQ		
		p.A37fs	p.E309_T310insSQ		
		p.G38fs	p.T310_Q312delinsK		
		p.G38fs	p.D320_N321ins16		
		p.A40fs	p.E309_L317dup		
		p.P43fs	p.Q312dup		
		p.P49fs	p.Q305_E316dup		
		p.E50fs	p.V314_L315ins13		
		p.G53fs	p.K313dup		
		p.G54fs	p.Q305_R306insQQ		
		p.I55fs	p.Q312_K313ins12		
		p.C56fs	p.K302_R306dup		
		p.E59fs	p.E309dup		
		p.E59fs	p.T318_S319insI		
		p.I62fs	p.E309delinsAQ		
		p.I62fs	p.Q312dup		
		p.I62fs	p.E329_Q330ins35		
		p.A66fs	p.T310_Q311insP		
		p.A66fs	p.E316_L317insR		
		p.Y67X	p.R297_V308dup		
		p.I68fs	p.R306_N307insRR		
		p.I68fs	p.K313dup		
		p.A72fs	p.K302_E309dup		
		p.F73fs	p.R306dup		
		p.F77fs	p.R306_Q311del		
		p.L78fs	p.R300_L324dupinsR		
		p.L78fs	p.V308_R327dup		
		p.L78fs	p.K313dup		
		p.A79fs	p.Q312dup		
		p.F82fs	p.K304_V308dup		
		p.Q83fs	p.R300_Q312dup		
		p.Q83fs	p.Q312_K313insE		
		p.Q83fs	p.K313dup		
		p.Q83fs	p.K313dup		
		p.H84fs	p.K302_K304dup		
		p.Q88fs	p.K304_Q305insL		
		p.A91fs	p.S319delinsRL		
		p.V95fs	p.V308I+Q312dup		
		p.G96fs	p.T310_Q311ins10		
		p.G96fs	p.K313dup		
		p.G96fs	p.L317_T318insM		
		p.T98fs	p.T310dup		
		p.G99fs	p.K304_Q305insL		
		p.G101fs	p.Q312dup		
		p.F106fs	p.D301_V308dup		
		p.F106fs	p.K302_R306dup		
		p.D107fs	p.S299_T318dup		
		p.D107X	p.D301del		
		p.D107fs	p.E309delinsGQ		
		p.Y108X	p.E309_T310insK		
		p.G110fs	p.Q312dup		

Clin (cont)		p.G110fs p.P112fs p.A113fs p.G114fs p.G114fs	p.Q312dup p.K304_Q305insL p.L315_E316insQ p.Q312dup p.K313dup		
	Non-classic DM	p.A44fs p.I68fs p.L78fs p.A79fs	p.A295P p.R297P p.R300P p.Q305P	Classic N + C-missense Classic N + C-missense Classic N + C-missense Classic N + C-missense	p30 + C-LOF
		p.H24fs p.H24fs p.G36fs p.G38fs p.G54fs p.E89fs (Hom) p.V95fs p.P121fs	p.Q209fs p.G223fs p.Y181X p.K313fs p.E167fs p.A238fs p.Q312X	Classic N + mid-frameshift Classic N + mid-frameshift Classic N + mid-frameshift Classic N + C-frameshift Classic N + mid-frameshift Classic N Classic N + mid-frameshift Classic N + C-frameshift	p30 + null
		p.L178fs p.R300_D301selinsQN (Hom) p.R306_V314dup p.Q312delinsPK p.K313dup p.K313dup p.K313dup (Hom) p.V314G (Hom) p.E316_L317insR p.E316_R325dup (Hom) p.L317Q (Hom) p.L317_K326dup (Hom) p.N321S (Hom)	p.K276R p.N321S p.R343fs p.R343fs p.S349fs p.R343fs	Mid-frameshift + C-missense Classic C Classic C + C-missense Classic C + C-frameshift Classic C + C-frameshift Classic C + C-frameshift Classic C C-missense Classic C + C-frameshift Classic C C-missense Classic C C-missense	C-LOF
M1	Classic DM	p.H24fs p.G54fs p.A66fs p.I68fs p.L78fs p.H84fs p.A91fs p.G114fs	p.Q311_Q312insL p.Q305_R306insQQ p.E316_L317insR p.R306_N307insRR p.E309_V328dup p.K302_K304dup p.S319delinsRL p.Q312dup	Classic N + C	p30 + C-LOF
	Non-classic DM	p.K313dup p.V314G (Hom)	p.R343fs	Classic C + C-frameshift C-missense	C-LOF
M2	Classic DM	p.P23fs p.H24fs p.H24fs p.H24fs p.A40fs p.P43fs p.E50fs p.I55fs p.E59fs p.Y67X p.A72fs p.A79fs p.Q88fs p.G96fs p.D107fs p.P112fs p.A113fs	p.N307_E309delinsK p.K304_Q305insL p.Q311_Q312insL p.K313Nins14 p.E309_L317dup p.Q312dup p.V314_L315ins13 p.Q312_K313ins12 p.E309dup p.R297_V308dup p.K302_E309dup p.Q312dup p.K304_Q305insL p.T310_Q311ins10 p.S299_T318dup p.K304_Q305insL p.L315_E316insQ	Classic N + C	p30 + C-LOF
	Non-classic DM	p.A44fs p.I68fs p.L78fs	p.A295P p.R297P p.R300P	Classic N + C-missense Classic N + C-missense Classic N + C-missense	p30 + C-LOF
		p.H24fs p.H24fs p.G36fs p.G38fs p.G54fs p.V95fs	p.Q209fs p.G233fs p.Y181X p.K313fs p.E167fs p.A238fs	Classic N + mid-frameshift Classic N + mid-frameshift Classic N + mid-nonsense Classic N + C-frameshift Classic N + mid-frameshift Classic N + mid-frameshift	p30 + null
		p.R300_D301delinsQN (Hom) p.E316_R325dup (Hom) p.N321S (Hom) p.L317Q (Hom) p.R306_V314dup	 p.N321S	Classic C Classic C C-missense C-missense Classic C + C-missense	C-LOF
	SM	p.P23fs p.H24fs p.P46fs p.E59X p.T60fs p.S61fs p.A66fs p.I68fs p.Q83fs		Classic N	p30 + WT

M2 (cont)		p.R300_D301delinsQY p.K313dup (x 2) p.R323del p.N356_C357del		Classic C	C-LOF + WT
		p.G122fs p.G123fs p.E166X p.D168fs p.P180fs p.Q215X p.L253fs p.S266fs p.K275fs		Mid-frameshift/nonsense	Null + WT
		p.N292fs p.N307fs		C-frameshift	Null + WT
		p.Y285S p.R289C p.V296E p.L331Q p.I341V		C-missense	C-LOF + WT
		p.P187_P189del p.H193_P196del p.P239_A240del		Mid-indel	UNK + WT
		p.P183Q p.P233R p.G242S (x 2) p.K276R		Mid-missense	UNK + WT

Null indicates mid-region or C-terminal mutants with a truncating frameshift or nonsense mutation.

Abbreviations: C, C-terminal mutation; Clin, Clinical cohort; Coh, cohort; DM, double *CEBPA* mutant; del, deletion; dupl, duplication; fs, frameshift; Hom, homozygous; indel, insertion/deletion; ins, insertion; LOF, loss-of-function; M1, Methylation cohort 1; M2, Methylation cohort 2; Mid, mid-region; N, N-terminal mutation; SM, single *CEBPA* mutant; UNK, unknown; WT, wild-type

Table S2. Demographic details of the 104 patients in the clinical analysis

Parameter	Classic *CEBPA*^{DM} (n=79)	Non-classic *CEBPA*^{DM} (n=25)	*P**
Median age, years (IQR)	35 (25 to 46)	47 (42 to 55_	0.001**
Median WBC, x10^9/L (IQR)	27.6 (10.5-67.4)	15.9 (7.7-92.9)	0.7**
Sex:			0.8
Male	44 (56%)	15 (60%)	
Female	35 (44%)	10 (40%)	
Trial:			0.7
AML10	16 (21%)	6 (24%)	
AML12	31 (39%)	11 (44%)	
AML15	32 (41%)	8 (32%)	
WHO Performance status:			0.1
0	55 (70%)	13 (52%)	
1	11 (14%)	8 (32%)	
2	12 (15%)	3 (12%)	
3	1 (1%)	1 (4%)	
AML type:			1.0
De novo	77 (97%)	25 (100%)	
Secondary	2 (3%)	0 (0%)	
Transplant status:			0.8
No transplant	48 (61%)	14 (56%)	
Allograft	19 (24%)	7 (28%)	
Autograft	9 (11%)	4 (16%)	
Other	3 (4%)	0 (0%)	
Stage of transplant:			0.7
First remission	18 (58%)	6 (55%)	
First relapse	1 (3%)	1 (9%)	
Second remission	12 (39%)	4 (36%)	
Cytogenetics:			0.05
FR	0 (0%)	1 (5%)	
IR	68 (100%)	17 (89%)	
NK	53 (78%)	12 (63%)	
AK	15 (22%)	5 (26%)	
AR	0 (0%)	1 (5%)	
No result	11	6 (24%)	
FLT3^{ITD}	8 (10%)	1 (4%)	0.7
NPM1^{MUT}	2 (3%)	1 (4%)	0.6
DNMT3A^{MUT}	2/76 (3%)	4 (16%)	0.03
IDH1^{MUT}	0/60 (0%)	1/21 (5%)	0.3
IDH2^{MUT}	1 (1%)	5 (20%)	0.003
WT1^{MUT}	7/71 (10%)	0/20 (0%)	0.3
TET2^{MUT}	3/30 (10%)	5/18 (28%)	0.1
GATA2^{MUT}	28/72 (39%)	4/23 (17%)	0.1

**P* values are Fishers exact test unless otherwise stated; **Wilcoxon rank sum test
Abbreviations: AK, intermediate-risk abnormal karyotype; AR, adverse risk; FR, favorable risk; IQR, interquartile range; IR, intermediate risk; NK, normal karyotype; WBC, white blood cell count

Table S3. The top 25 most differentially methylated sites between *CEBPA*$^{\text{Classic-DM}}$ and *CEBPA*$^{\text{WT}}$ samples. The probes were selected by Wilcoxon tests and the median difference between the samples (mean rank of both parameters) and are ordered by their median beta value in *CEBPA*$^{\text{Classic-DM}}$ samples.

Genomic Location				Gene			Signature Details					
Probe ID	Chr	Position	Dist. to TSS	ENSEMBL Gene ID	Symbol	Description	In CpG Island	WT Median β	Classic DM Median β	Wilcoxon P-value	Median Δ	Rank
cg21237418	17	24069170	-157	ENSG00000109113	RAB34	RAB34, member RAS oncogene family	FALSE	0.153	0.928	4.91E-07	0.775	2
cg14338887	6	43036478	0	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.170	0.918	3.93E-06	0.748	3
cg17186163	10	44794323	7	ENSG00000165507	C10orf10	Chromosome 10 open reading frame 10	FALSE	0.155	0.907	4.09E-08	0.752	1
cg24101359	6	43036473	5	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.199	0.904	1.89E-05	0.705	20
cg13105904	14	23969884	-903	ENSG00000100441	KHNYN	KH and NYN domain containing	TRUE	0.306	0.889	5.56E-06	0.583	10
cg01274660	7	100303561	-675	ENSG00000087077	TRIP6	Thyroid hormone receptor interactor 6	FALSE	0.317	0.888	1.80E-05	0.572	25
cg04355435	1	43508877	117	ENSG00000179178	TMEM125	Transmembrane protein 125	FALSE	0.349	0.879	7.69E-06	0.531	22
cg10056627	6	43036751	-273	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.249	0.877	1.42E-05	0.629	16
cg27588902	6	43036129	349	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.265	0.854	2.74E-06	0.589	4
cg25651505	2	85665534	-492	ENSG00000168899	VAMP5	Vesicle associated membrane protein 5	TRUE	0.347	0.837	2.74E-06	0.489	18
cg23696834	6	43036323	155	ENSG00000124713	GNMT	Glycine N-methyltransferase	TRUE	0.102	0.822	3.93E-06	0.720	5
cg24081819	8	27404857	-295	ENSG00000120915	EPHX2	Epoxide hydrolase 2, cytoplasmic	TRUE	0.243	0.817	3.93E-06	0.574	9
cg08965235	11	65081734	541	ENSG00000168056	LTBP3	Latent TGF beta binding protein 3	TRUE	0.266	0.804	2.74E-06	0.538	11
cg16068833	1	26517102	-104	ENSG00000169442	CD52	CD52 molecule	FALSE	0.227	0.763	1.06E-05	0.536	24
cg19764555	11	62071695	-787	ENSG00000124942	AHNAK	AHNAK nucleoprotein	TRUE	0.272	0.763	2.74E-06	0.491	17
cg00350296	11	65841417	-343	ENSG00000174807	CD248	CD248 molecule, endosialin	FALSE	0.253	0.738	2.74E-06	0.485	19
cg10798171	7	8268826	-59	ENSG00000003147	ICA1	Islet cell autoantigen 1	TRUE	0.249	0.715	4.09E-08	0.466	13
cg15032239	15	20443395	709	ENSG00000068793	CYFIP1	Cytoplasmic FMR1 interacting protein 1	TRUE	0.195	0.708	4.09E-08	0.513	6
cg16155382	1	24518778	-135	ENSG00000158055	GRHL3	Grainyhead-like transcription factor 3	FALSE	0.101	0.678	3.69E-06	0.577	8
cg13490971	5	141468305	203	ENSG00000131507	NDFIP1	Nedd4 family interacting protein 1	TRUE	0.203	0.653	4.09E-08	0.450	14
cg21697134	17	78287128	-331	ENSG00000167363	FN3K	Fructosamine 3 kinase	FALSE	0.090	0.614	1.04E-07	0.524	7
cg08897388	6	112682398	44	ENSG00000112769	LAMA4	Laminin subunit alpha 4	TRUE	0.135	0.575	1.84E-06	0.440	23
cg12417466	3	35658823	30	ENSG00000172995	ARPP21	CAMP regulated phosphoprotein 21kDa	FALSE	0.720	0.179	5.56E-06	-0.541	15
cg05615150	3	35658819	34	ENSG00000172995	ARPP21	CAMP regulated phosphoprotein 21kDa	FALSE	0.668	0.113	5.56E-06	-0.555	12
cg18920397	1	159032429	123	ENSG00000122224	LY9	Lymphocyte antigen 9	FALSE	0.625	0.071	1.06E-05	-0.554	21

Table S4. Details of primers and conditions for methylation-specific PCR and pyrosequencing assays.

	Gene	Primer	Primer Sequence	Size (bp)	Annealing Temp (°C)
Methylation-specific PCR for bisulfite conversion	*HLA-B* converted	F	5'-TTTTAAGTTTTATTTTTGTGGGGTA-3'	300	Touchdown
		R	5'-AAATCCCAACTAATAACTATTTTTCAA-3'		(see text)
	HLA-B unconverted	F	5'-CCCAAAGTCCACTAACATTAGAA-3'	464	Touchdown
		R	5'-GCTGAGAAAATAGCCTCAGAATA-3'		(see text)
Pyrosequencing assays for array validation	*SOCS2*	F (biotinylated)	5'-AGGTGGGAAGTAAAGAATAAGATGGA-3'	128	62
		R	5'-CCAAACCTAAATCCCTAAAAAACCACTTT-3'		
		Sequencing	5'-CCTAAAAAACCACTTTCCT-3'		
	WNT2	F	5'-GTGTATGAAATGATGGTAAGAGATGTT-3'	246	62
		R (biotinylated)	5'-ATACATAATAATCTCCTTATCCCCTAACC-3'		
		Sequencing	5'-GGGAAGGGGGAATATYGTTGTATG-3'		
	PRF1 A	F	5'-AGTAGGGTTATTTTTTTGTTTTTGATGT-3'	150	60
		R (biotinylated)	5'-CCTACCAATCCACACTACTAATACA-3'		
		Sequencing	5'-GTTATTTTTTTGTTTTTGATGTATA-3'		
	PRF1 B	F	5'-TAGGAAGTGTTGTGATTTATAAGATAAG-3'	163	60
		R (biotinylated)	5'CTTTAATATCAACACTTACAAAACCTTAA-3'		
		Sequencing	5'-TAAGATAAGATATTTGGGTTA-3'		
Pyrosequencing assays for good-risk patients	*LY9*	F (biotinylated)	5'-TGTTTTAGAGGGAGGGTTGTTTATA-3'	100	58
		R	5'-AATCACAAATAAAACCCTAAATAAACTTA-3'		
		Sequencing	5'-TAAAACCTCTACCTACC-3'		
	KHNYN	F (biotinylated)	5'-GGGTTTTTTAGTTGTAGTTAGATGTG-3'	192	60
		R	5'-ACTAAAAACAACAACCATACCTAC-3'		
		Sequencing	5'-ACCCCATATAAAACCCATCTTC-3'		
	VAMP5	F	5'-GTGTTYGTTTATTAGGTAGAGGTGTTA-3'	281	59
		R (biotinylated)	5'-CCCRCCTAAACCCTCACCATC-3'		
		Sequencing	5'- GTTTATYGTTTTYGATTTGATTTGG-3'		

SUPPLEMENTARY FIGURES

Figure S1. Flowchart showing the patients investigated in the different analyses performed.

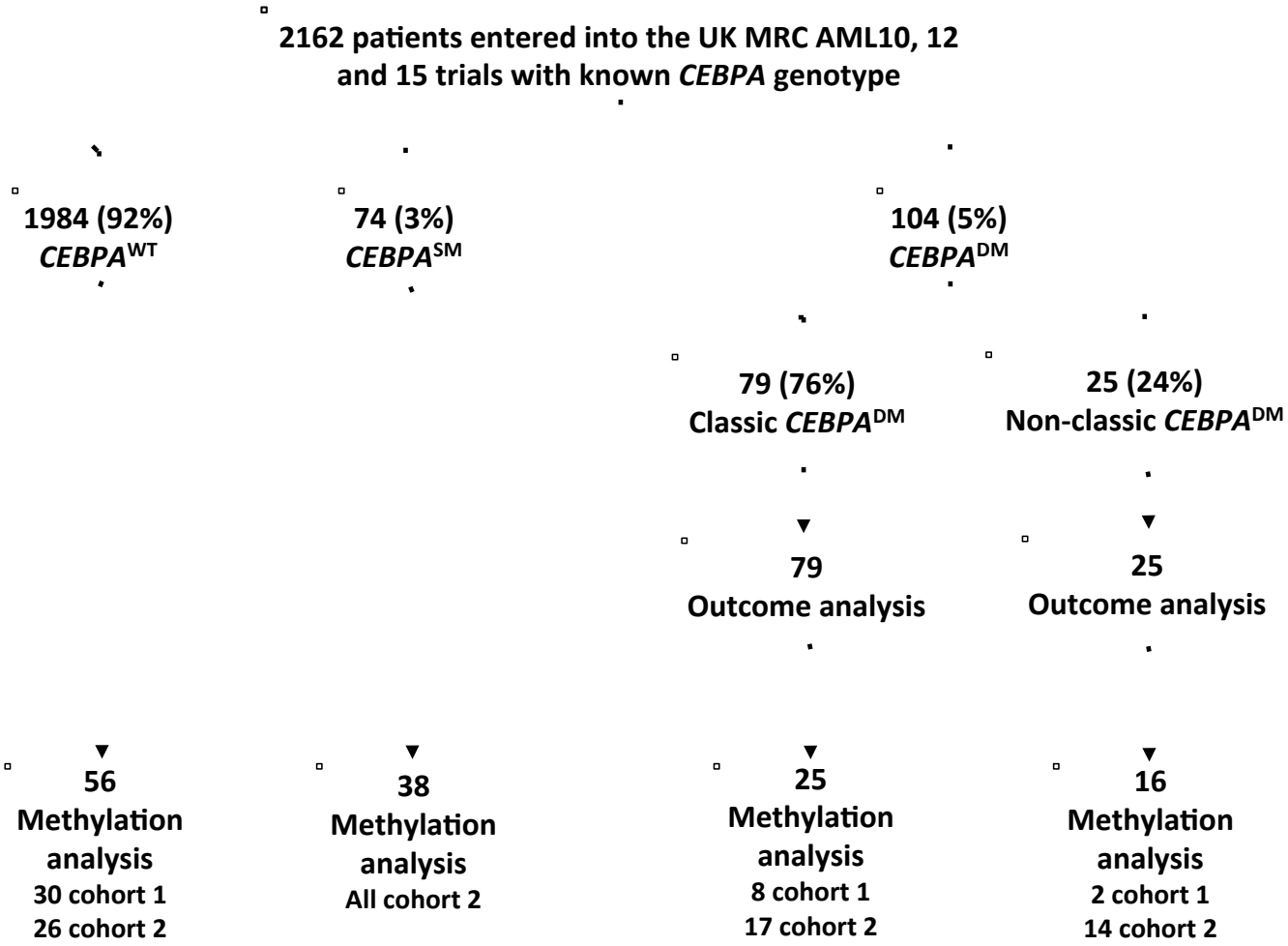


Figure S2. Comparison of methylation quantification values obtained at four CpG sites investigated on the arrays and by pyrosequencing using the preliminary set of 40 samples. (a-d) β values obtained from the methylation array versus % methylation by pyrosequencing. (e-h) Bland-Altman plots showing the mean result for each sample compared with the difference in values between the two assays. The dotted line indicates the 95% confidence intervals for each probe. Details of PCR primers and assay conditions are given in Table S4.

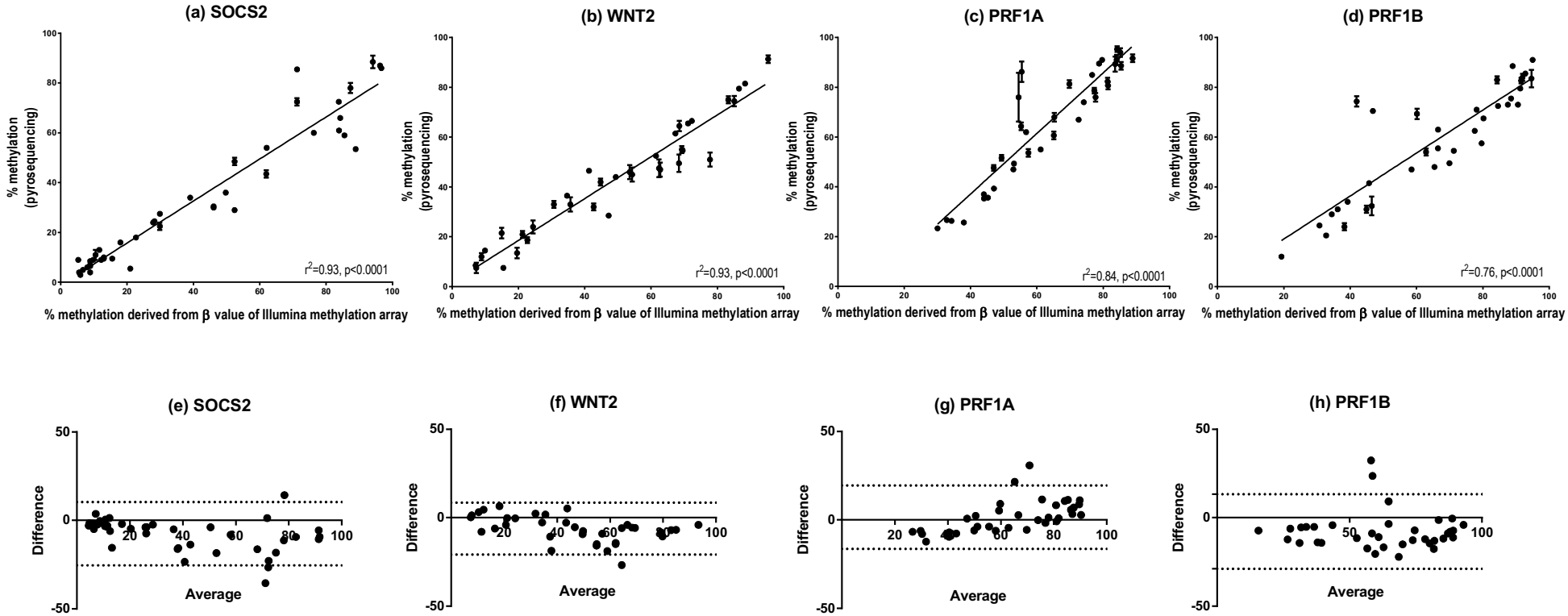


Figure S3. Unsupervised cluster analysis of the follow-up cohort. Each column represents a patient. Genotype is given in the upper panel. Samples were clustered based on their methylation levels at 157,797 variable probes and the heatmap in the middle panel shows the variable CpG probes located within CGIs. The latter were used to calculate the mean % CpG methylation shown in the lower panel; red and blue bars indicate a predominantly hyper- or hypo-methylated profile respectively.

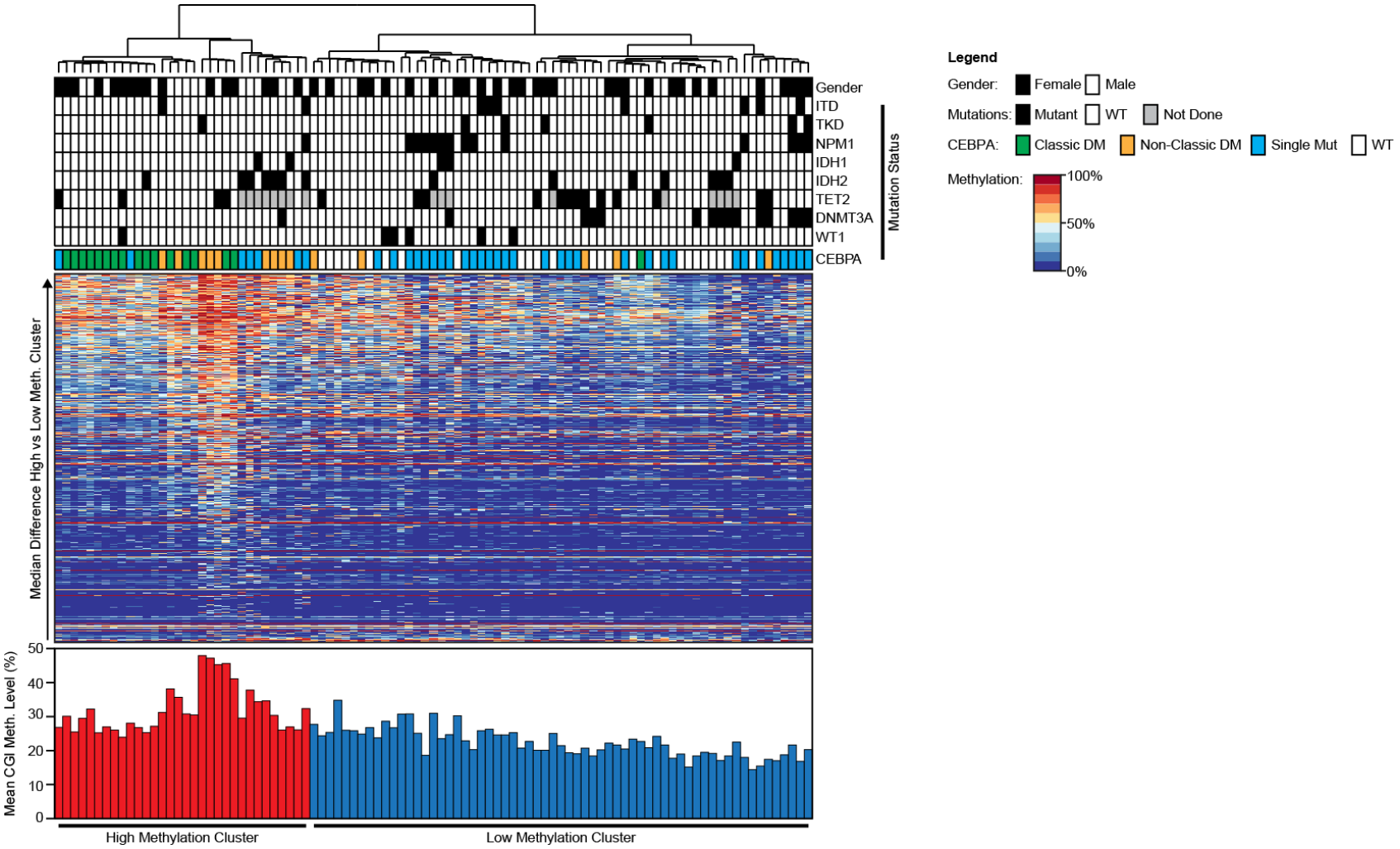


Figure S4. Methylation levels in other good-risk groups compared to the *CEBPA*$^{\text{Classic-DM}}$ cases at three differentially methylated CpG sites. The 24 *CEBPA*$^{\text{Classic-DM}}$ (excluding the outlier) and 56 *CEBPA*$^{\text{WT}}$ results were beta values from the arrays; results for the three comparative groups were obtained by pyrosequencing. Significance refers to difference from the *CEBPA*$^{\text{Classic-DM}}$ group (**P*≤0.05, ***P*≤0.01, ****P*≤0.001)